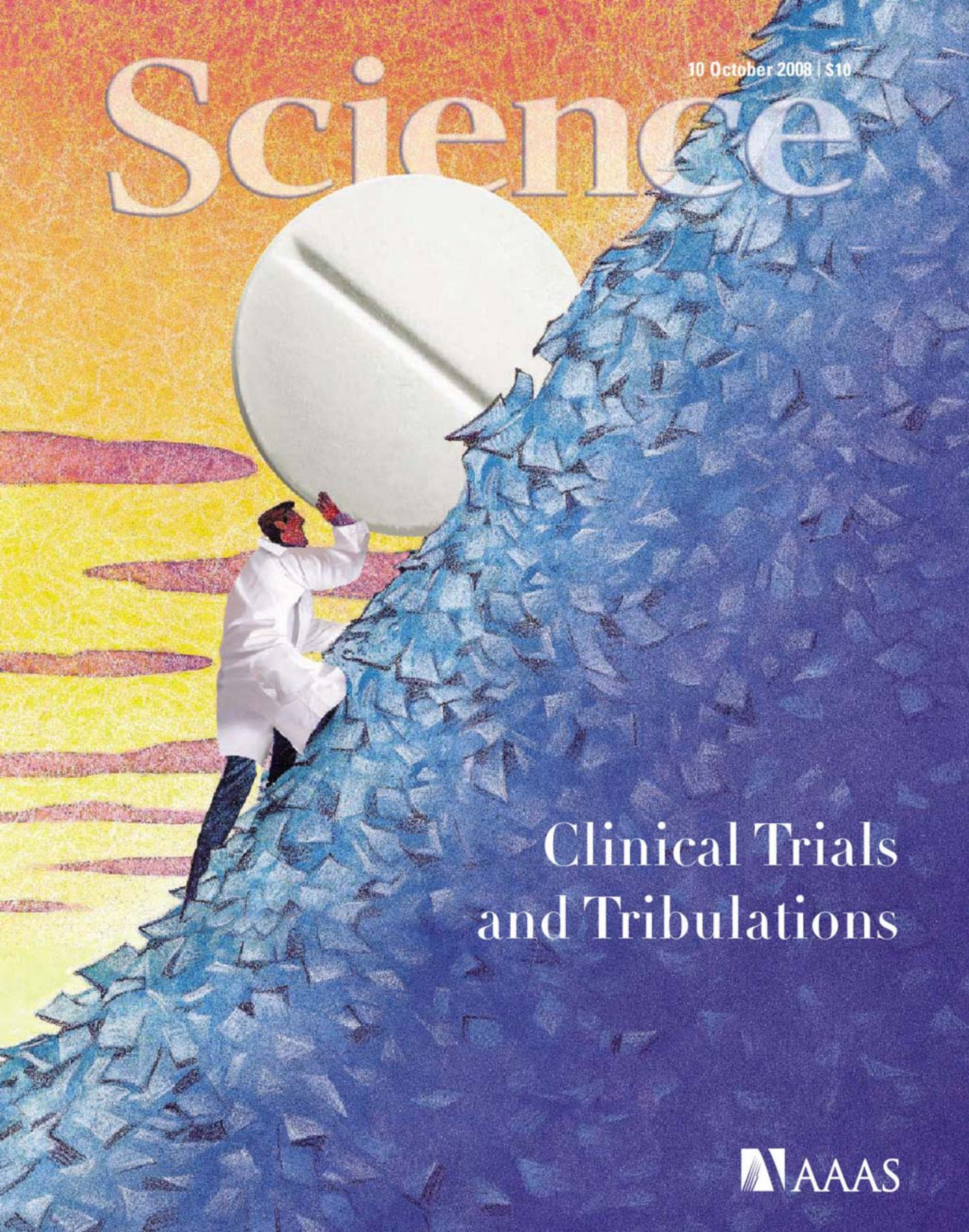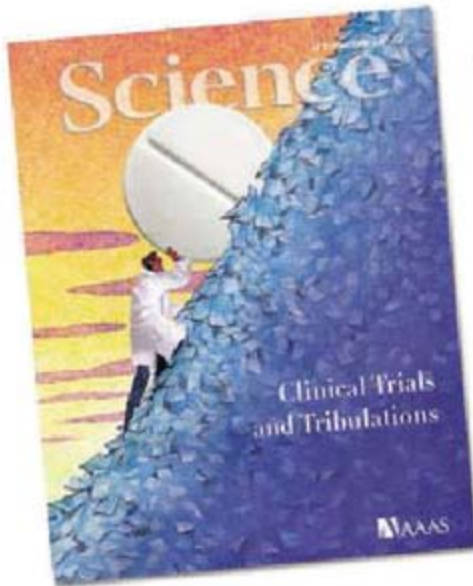# Science

Clinical Trials and Tribulations

AAAS

COVER

Steep terrain: To move a therapy from the research lab to the doctor's office requires a huge investment in clinical trials, which are growing more costly and more complex every year. See the special section beginning on page 209.

Photo illustration: Kelly Buckheit Krause (images: Getty Images; Jupiter Images)

DEPARTMENTS

EDITORIAL

SPECIAL SECTION

Clinical Trials

INTRODUCTION

NEWS

For related online content, see page 159 or go to www.sciencemag.org/clinicaltrials/

NEWS OF THE WEEK

NEWS FOCUS

SCIENCE EXPRESS

LETTERS

BOOKS *ET AL.*

POLICY FORUM

PERSPECTIVES

BREVIA

RESEARCH ARTICLES

REPORTS

225

Science

REPORTS CONTINUED...

246

Printed on
30% post-consumer
recycled paper.

SCIENCE NOW

www.sciencenow.org

HIGHLIGHTS FROM OUR DAILY NEWS COVERAGE

Ig Nobels Honor Studies of Lap Dancing, Soft Drink–Based Contraception
Comic take on Nobels awarded to 10 research teams, amid flying paper airplanes.

Blood Test for Mom Picks Up Down Syndrome in Fetus
New approach offers hope for cheaper, less invasive diagnosis.

Caving for Seismic History
Stalagmites may serve as record of past earthquakes.

Understanding gum disease.

SCIENCE SIGNALING

www.sciencesignaling.org

THE SIGNAL TRANSDUCTION KNOWLEDGE ENVIRONMENT

PODCAST
N. R. Gough and J. F. Foley
The bacteria that cause periodontitis evade destruction by promoting crosstalk between two receptors that regulate the immune response.

PRESENTATION: The Endothelial Glycocalyx—A Mechano-Sensor and -Transducer
J. M. Tarbell and E. E. Ebong
A complex extracellular network of proteoglycans communicates mechanical stress to endothelial cells.

ST NETWATCH: Cellerator
This plug in for Mathematica allows users to model interactions between components of a signal transduction network or between multiple networks; in Modeling Tools.

ST NETWATCH: Clapham Lab
Find a wealth of information about TRP channels and calcium signaling from this lab at Harvard Medical School; in Labs and People.

ST NETWATCH: GenMAPP
Make sense of array data by using this application to map gene expression data onto cellular pathways; in Bioinformatics Resources.

SCIENCE ONLINE FEATURE

VIDEO: Pediatric Medicine—Prescribing Drugs "Off-Label"
A progress report on regulating and testing medicines for kids.
>> *Clinical Trials section* *or go to*
www.sciencemag.org/clinicaltrials/

Planning your research.

SCIENCE CAREERS

www.sciencecareers.org/career_development

FREE CAREER RESOURCES FOR SCIENTISTS

Maximizing Productivity and Recognition, Part 3: Developing a Research Plan
S. Pfirman, R. E. Bell, P. J. Culligan, P. Balsam, J. D. Laird
To set your research on a clear course, you need a well-conceived but flexible plan.

Opportunities: The Political Scientist
P. Fiske
Not understanding political elements of the scientific culture can reduce your effectiveness.

Transferring Skills to Tech Transfer
L. Laursen
Technology transfer departments can bridge the gap between science and business.

Science Careers Blog
Science *Careers Staff*
Get advice, opinions, news, funding opportunities, and links to other career-related resources.

SCIENCE PODCAST

www.sciencemag.org/multimedia/podcast

FREE WEEKLY SHOW

Download the 10 October *Science* Podcast to hear about digital blueprints of vertebrate development, a single-species ecosystem deep within the Earth, clinical trials gridlock, and more.

CREDITS: (SCIENCE CAREERS) PHOTODISC; (SCIENCE SIGNALING) JUPITER IMAGES; (SCIENCE ONLINE FEATURE) JUPITER IMAGES

Separate individual or institutional subscriptions to these products may be required for full-text access.

ARMOR PLANTING

Sticklebacks residing in marine settings differ from their freshwater cousins by possessing more protective lateral plates. How does the colonization of fresh water from marine habitats result in this loss of armor plating along the body? **Barrett** *et al.* (p. 255, published online 28 August; see the Perspective by **Cresko**) replicated sticklebacks' natural colonizations from marine to freshwater environments and measured the frequency of alleles affecting plate number among the progeny. Selection against lateral plate development was observed in freshwater habitats but the selection gradient changed over time, with larger (older) fish being more likely to possess reduced armor.

About Africa

Tropical Africa is an important locus of the global hydrological cycle, and home to a large number of people, but records of how rainfall and temperature have varied in the past are too few and too short to provide a good understanding of what has controlled them over the geologically recent past. **Tierney** *et al.* (p. 252, published online 11 September) constructed records of temperature and precipitation from sediments from Lake Tanganyika for the past 60,000 years in order to help to clarify what factors drive climate in the region. Clear links were observed to the Northern Hemisphere and the Indian Ocean, and, surprisingly, changes in the position of the Intertropical Convergence Zone were not the dominant control on rainfall variations. Southeast African climate is thus driven by a complicated, incompletely understood, array of forces.

Triple Cooling

In contrast to the cold atom regimes of matter vigorously studied over the past decade, a sample of polar molecules cooled to a near stand-still would offer a complex added range of intermolecular interactions. A stumbling block in generating such a sample, however, has been the need to dissipate internal energy tied up in molecular vibrations and rotations. **Ni** *et al.* (p. 231, published online 18 September; see the Perspective by **Gould**) surmount this challenge by using a coherent Raman process, induced by irradiation at two different laser frequencies. The technique yields a translationally ultracold sample of KRb heterodiatomics in the bottom vibrational and rotational levels of the lowest-energy triplet electronic state.

Fluorescing in Sequence

The capacity of DNA to resist photodamage by rapidly dissipating the energy from ultraviolet light absorption is a key feature of long-term genetic stability. Many model studies have probed the underlying dynamics responsible for this dissipation, but **Schwalb and Temps** (p. 243) now find that structural simplifications in such studies can be risky. Comparisons of picosecond fluorescent lifetimes for a range of oligonucleotides, both single- and double-stranded, revealed a complex dependence on base sequence. Substitution of four guanine-cytosine pairs into a 20-membered adenine-thymine strand halved the fluorescent lifetime, highlighting a general trend in which mixed sequences relax more rapidly.

Sticky Stuff

For sticking an object to a wall, an ideal adhesive would have a strong shear adhesive force to counteract gravity, and much lower normal adhesive force to allow for easy removal and reattachment. Significant efforts have been made to mimic the foot of the gecko, because the animal is able to climb almost any surface. **Qu** *et al.* (p. 238) demonstrate that vertically aligned carbon nanotube arrays can achieve very high dry shear strength on a variety of surfaces. A thin disordered nanotube layer on top of the aligned nanotube array leads to a significant adhesion anisotropy with shear strengths far exceeding the normal. The shear adhesive force achieved is greater than that of the gecko, creating a strong adhesive affect that can still be moved from one surface to another.

Monitoring Mammalian Conservation Status

The development of appropriate conservation strategies for global mammalian diversity requires knowledge of effective strategies and

details of the current status of at-risk species. **Schipper** *et al.* (p. 225) present the key findings of the Global Mammal Assessment, a project involving more than 1700 collaborators in 130 countries and entailing the compilation of standardized data on the distribution and conservation status of all 5487 mammal species (including marine species). The database, made freely available by the International Union for the Conservation of Nature, allows a thorough description of patterns of diversity, threat, and knowledge for all species. Even though apparently quite different patterns were found for marine and land species, the results suggest that they may be underpinned by similar mechanisms. The results also reveal the bleak conservation status of the world's mammals, with one in four land species and one in three marine species under threat. Land species are mainly threatened by habitat loss and by harvesting, while marine species are mainly accidental victims of fishing practices and pollution.

Seising Up Soufrière

The dynamics of volcanic eruptions coupled with observations of deformation of the ground surface and seismic and geochemical data can provide a view of the plumbing system feeding the volcano. **Elsworth** *et al.* (p. 246) integrate such data for the eruption of the Soufrière Hills volcano, Montserrat, which has been ongoing for more than 12 years. Three major eruption cycles have occurred. Correlation of ground inflation and deflation with eruption style and intensity show that magma from the mantle or base of the crust is filling a mid-crustal reservoir continuously, which fills a compliant subvolcanic reservoir. The data also suggest that the eruption is now nearing its end.

Hot, Deep, and Alone

Organisms can thrive many kilometers below the Earth's surface in the absence of sunlight and oxygen. **Chivian** *et al.* (p. 275) describe the assembly and annotation of a genome from a bacterium in a deep aquifer in a South African gold mine. The genomic analyses and annotation suggest it has the capacity to be a motile, sporulating, sulfate reducing, chemoautotrophic thermophile with a near clonal population structure. This single-species community has a life-style sustained by the hydrolysis of water, which generates hydrogen, as the reductant for energy, and hydrogen peroxide, which oxidizes iron sulfide minerals to sulfate.

A Century in Yosemite

In the early–20th century, Joseph Grinnell, a founder of ecology, and his colleagues surveyed small-mammal diversity across an elevational transect of 60 to 3300 meters in what is now the Yosemite National Park in California, USA. Their surveys were highly detailed and well-documented, and the data, backed by specimens, field notes, and photographs, have been retained in the Museum of Vertebrate Zoology. Nearly 100 years later, **Moritz** *et al.* (p. 261; see the Perspective by **Svenning and Condit**) resurveyed small-mammal diversity across the Yosemite transect. The results reveal a clear community-scale response to climate warming, with low-elevation species expanding upward and high-elevation species contracting. This century-scale record provides hard evidence for substantial changes in response to past climate change, including the range collapse of some high elevation species. Species diversity has, however, been retained, despite range fluctuations, which suggests the importance of protected landscapes for retaining diversity through migratory responses to climate change.

Promoting Toxin Processing

Bacterial toxins are often produced as inactive precursors that are activated when they enter eukaryotic cells. One family of such toxins, the Multifunctional Autoprocessing Toxins (MARTX), which includes *Vibrio cholerae* RTX, are activated by their own cysteine protease domain (CPD) and, recently, the small molecule inositol hexaphosphate ($InsP_6$) has been shown to stimulate this autoproteolysis. **Lupardus** *et al.* (p. 265) have determined the crystal structure of the *Vibrio cholerae* RTX CPD bound to $InsP_6$. Combining the structural data with biochemical and kinetic analysis of CPD mutants suggests that $InsP_6$ binding induces an allosteric switch that exposes the CPD active site and leads to autoprocessing.

CREDIT: THANYA SUWANSAWAD

Kai Simons is president of the European Life Scientist Organization and is at the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden, Germany.

The Misused Impact Factor

RESEARCH PAPERS FROM ALL OVER THE WORLD ARE PUBLISHED IN THOUSANDS OF SCIENCE journals every year. The quality of these papers clearly has to be evaluated, not only to determine their accuracy and contribution to fields of research, but also to help make informed decisions about rewarding scientists with funding and appointments to research positions. One measure often used to determine the quality of a paper is the so-called "impact factor" of the journal in which it was published. This citation-based metric is meant to rank scientific journals, but there have been numerous criticisms over the years of its use as a measure of the quality of individual research papers. Still, this misuse persists. Why?

The annual release of newly calculated impact factors has become a big event. Each year, Thomson Reuters extracts the references from more than 9000 journals and calculates the impact factor for each journal by taking the number of citations to articles published by the journal in the previous 2 years and dividing this by the number of articles published by the journal during those same years. The top-ranked journals in biology, for example, have impact factors of 35 to 40 citations per article. Publishers and editors celebrate any increase, whereas a decrease can send them into a huddle to figure out ways to boost their ranking.

This algorithm is not a simple measure of quality, and a major criticism is that the calculation can be manipulated by journals. For example, review articles are more frequently cited than primary research papers, so reviews increase a journal's impact factor. In many journals, the number of reviews has therefore increased dramatically, and in new trendy areas, the number of reviews sometimes approaches that of primary research papers in the field. Many journals now publish commentary-type articles, which are also counted in the numerator. Amazingly, the calculation also includes citations to retracted papers, not to mention articles containing falsified data (not yet retracted) that continue to be cited. The denominator, on the other hand, includes only primary research papers and reviews.

Why does impact factor matter so much to the scientific community, further inflating its importance? Unfortunately, these numbers are increasingly used to assess individual papers, scientists, and institutions. Thus, governments are using bibliometrics based on journal impact factors to rank universities and research institutions. Hiring, faculty-promoting, and grant-awarding committees can use a journal's impact factor as a convenient shortcut to rate a paper without reading it. Such practices compel scientists to submit their papers to journals at the top of the impact factor ladder, circulating progressively through journals further down the rungs when they are rejected. This not only wastes time for editors and those who peer-review the papers, but it is discouraging for scientists, regardless of the stage of their career.

Fortunately, some new practices are being attempted. The Howard Hughes Medical Institute is now innovating their evaluating practices by considering only a subset of publications chosen by a scientist for the review board to evaluate carefully. More institutions should determine quality in this manner.

At the same time, some publishers are exploring new practices. For instance, *PLoS One*, one of the journals published by the Public Library of Science, evaluates papers only for technical accuracy and not subjectively for their potential impact on a field. The European Molecular Biology Organization is also rethinking its publication activities, with the goal of providing a means to publish peer-reviewed scientific data without the demotivating practices that scientists often encounter today.

There are no numerical shortcuts for evaluating research quality. What counts is the quality of a scientist's work wherever it is published. That quality is ultimately judged by scientists, raising the issue of the process by which scientists review each others' research. However, unless publishers, scientists, and institutions make serious efforts to change how the impact of each individual scientist's work is determined, the scientific community will be doomed to live by the numerically driven motto, "survival by your impact factors."

– Kai Simons

NEUROSCIENCE
Staggered Starts

If you are a small animal with many predators in your environment, you'd better react quickly to the earliest sign of danger. Startle reflexes have evolved as one way to deal with such situations. Input from sensory organs is rapidly transmitted to efferent systems, usually the muscles that are involved in flight responses. One such reflex is the backward tailflip response in crayfish. The coincident arrival of synaptic inputs is crucial because only when enough synapses fire simultaneously will they be able to activate the giant fiber system; however, the medial giant fibers receive input from widely differing distances along the crayfish antennules. Mellon and Christison-Lagay found that sensillar axons have precisely calibrated conduction velocities. Axon diameters near the flagellum tip were very small, and axon diameters increased continuously toward more proximal locations along the flagellar axis. In nonmyelinated nerve fibers, conduction velocity is proportional to the square root of the axon diameter. This linear adjustment of axonal conduction velocity thus ensures that a train of action potentials from an array of feathered sensilla on the first antennae arrives simultaneously in the brain and can drive the medial giant fibers to produce the tailflip response. — PRS

Proc. Natl. Acad. Sci. U.S.A. **105**, 14626 (2008).

CLIMATE SCIENCE
What Have We Done?

Global climate is warming rapidly, and has been for more than 150 years, since around the start of the Industrial Revolution. Climate models suggest that most of the rise is due to greenhouse gas emissions, but the accuracy of the models is not entirely certain, and there have been numerous high-profile disagreements about their credibility. An alternate way to estimate the magnitude of human influence on global temperatures is to look at the observational record. Lean and Rind found that only four factors—ENSO (El Niño–Southern Oscillation), volcanic activity, solar activity, and anthropogenic forcing by greenhouse gases—are required to explain 76% of the variance in the temperature records. Furthermore 90% or more of the warming trend of the past 100 years can be explained by invoking anthropogenic effects, and solar forcing can explain a negligible percentage of the rise in temperature over the past 25 years. Finally, the zonal temperature response to natural and anthropogenic forcing does not increase rapidly with latitude from mid- to high latitudes, as it does in models, and anthropogenic warming effects are more pronounced in the latitudes between 45°S and 50°N than at higher latitudes. — HJS

Geophys. Res. Lett. **35**, L18701 (2008).

CELL BIOLOGY
Decisions, Decisions

Precursor cells begin life faced with some important decisions: when and where to divide, and what sort of cell to become. These decisions can be influenced chemically by growth factors that induce signal transduction pathways within the cell and regulate the transcription of specific genes. Oligodendrocytes make myelin that insulates axons; they arise from precursor cells that migrate and proliferate along the axons before differentiating into mature myelin-forming cells. Rosenberg *et al.* show that the physical environment of the oligodendrocyte precursor cell acts as a differentiation factor and influences these cell fate decisions. Differentiation was dependent on the packing density of the precursor cells, suggesting that lateral compression of the cells can act as a mechanical signal that activates gene expression programs. — HP*

Proc. Natl. Acad. Sci. U.S.A. **105**, 14662 (2008).

CHEMISTRY
Spinning Soy into Gold

Many routes exist for making metal nanoparticles but often involve non–water-soluble and/or potentially toxic compounds, a particular concern in preparations for in vivo applications. Shukla *et al.* prepared gold nanoparticles from aqueous solutions of sodium tetrachloroaurate mixed with soybean extracts. Soybeans contain a wealth of phytochemicals, including proteins, carbohydrates, fats, and isoflavones. Gold nanoparticles formed over the course of a day by direct interaction with soybeans in solution, but better results were achieved using extracts from presoaked soybeans to afford time for the proteins to dissolve. Both low- and high-molecular-weight protein fractions were able to reduce the gold ions, but stabilization of the particles was observed only with the higher-molecular-weight fraction, which effectively coated the particles

Field of gold?

*Helen Pickersgill and Chris Surridge are locum editors in *Science*'s editorial department.

CREDITS (TOP TO BOTTOM): JUPITERIMAGES

and prevented aggregation. Particles were stable in salt, histidine, and bovine and human serum albumin solutions and did not show cytotoxicity when tested with human fibroblast cells. Thus, plant matter may be an effective route to making biomedically friendly metal nanoparticles. — MSL

Small 4, 1425 (2008).

ECOLOGY
The Me Generation

Despite conservation efforts to ensure its survival, an island bird species has remained endangered because of its own social behaviors. The Seychelles magpie robin, whose population dwindled to just seven breeding pairs in 1988, has been the subject of intense conservation efforts for more than three decades. Success has been modest, with reintroductions within the Seychelles archipelago boosting

Seychelles magpie robin.

the population size to almost 150 individuals in all. López-Sepulcre *et al.* show that recovery has been slower than expected because of competition within territorial social groups consisting of a dominant breeding pair and several subordinate individuals, whose reproduction is postponed until one of the dominant pair dies or is competitively ousted. Competitive interactions for dominance within the group reduce the group's reproductive output. Thus, behavior that is advantageous for an individual's fitness (and evolutionary success) can be detrimental to that of the group. Simulations suggest that recovery of the population would have been at least one-third more rapid in the absence of competitive interactions, highlighting the need to take social behavior into account in species recovery plans. — AMS

J. Anim. Ecol., 10.1111/
j.1365-2656.2008.01475.x (2008).

IMMUNOLOGY
Hard Graft Made Easy

Type 1 diabetes is an autoimmune disease that destroys insulin-producing β cells, situated in the pancreas within the islets of Langerhans, leaving sufferers dependent on regular injections of insulin to control their blood glucose levels. An attractive treatment would be the transplantation

of islets from healthy donors but, as with all organ transplants, there is the risk of rejection and a need for long-term suppression of the recipient's immune system, leaving the person prey to opportunistic infections. Luo *et al.* have developed a method to make diabetic mice tolerant to islet grafts by injecting them once 1 week before transplantation and again 1 day afterward with donor spleen cells, which were first treated with the chemical crosslinker 1-ethyl-3-(3'-dimethylaminopropyl)-carbodiimide. Antigen-presenting cells from the donor spleen induced the down-regulation of the host effector T cells that would otherwise orchestrate graft rejection, and encouraged regulatory T cells to provide long-term tolerance to the transplants. Islet cells grafted into diabetic mice produced insulin for several months, and grafts could be replaced without additional treatment, as long as the new islets came from the same original donor. This approach depended on the exact timing and size of fixed-cell injections but, if a similar protocol can be established for humans, it could provide a simple and effective therapy for a very common condition. — CS*

Proc. Natl. Acad. Sci. U.S.A.
105, 14527 (2008).

DEVELOPMENTAL BIOLOGY
Alternative Route to Male Killing

Wildly distorted sex ratios in certain insects— where as few as 1 male per 100 offspring are produced, dramatically altering patterns of mate competition—are caused by "male-killing" bacteria. Male killing in the parasitic wasp *Nasonia vitripennis*, in which males develop from unfertilized eggs by parthenogenesis, is caused by the bacterium *Arsenophonus nasoniae*.

Ferree *et al.* show that in infected *Nasonia* females, unfertilized eggs die during early embryogenesis because of a lack of centrosome activity and the consequent disarray of cell division. In diploid species, centrosomes are generally provided by the sperm. In *Nasonia*, which have haploid eggs, the centrosomes are derived from "accessory nuclei." Accessory nuclei, vesicular organelles that are formed from the oocyte nuclear membrane that sequester centrosome components, are intact in both infected and uninfected wasps, suggesting that *Arsenophonus* interferes with events downstream of their formation. Indeed, haploid male eggs can be rescued by the delivery of paternal centrosomes. Thus, *Arsenophonus* interferes with the development of unfertilized eggs rather than targeting maleness. — GR

Curr. Biol. **18**, 1409 (2008).

AAAS
ADVANCING SCIENCE, SERVING SOCIETY

Big Brother Is Watching You—and Your Food

Want to grab lunch at the Restaurant of the Future? Fine—if you're okay with your entire visit being videotaped and analyzed by scientists. The new restaurant, set up by Wageningen University in the Netherlands and three companies specializing in food, software, and kitchens, is actually a laboratory where researchers can study eating and drinking behavior. Casino-like ceiling cameras track your every move, and a hidden scale surreptitiously weighs you before you dig in. Lighting, furniture, and even the ambient odor can be changed to create different moods.

The place, which officially opened on 4 October, also boasts a sensory research lab in which researchers can measure how people look at, choose, chew, and swallow food. The information can help companies refine their products and lead to new strategies to promote healthier eating, says René Koster, director of the "Restaurant of the Future" research foundation. The setup is an excellent way to study consumer behavior under natural circumstances, says food psychologist Brian Wansink of Cornell University, who runs a similar research restaurant called the Spice Box in Urbana, Illinois. "I've gotten tremendous value out of it," Wansink says—including much of the material in his book *Mindless Eating: Why We Eat More Than We Think.*

AARP Lauds NIH

The National Institutes of Health (NIH) has been engaging in a lot of hand-wringing about the aging of the biomedical research workforce. But it turns out not everybody is unhappy: AARP (formerly the American Association of Retired Persons) has just named NIH one of the best employers for workers over 50. The first federal agency to make the list, NIH ranks 11th after employers such as Cornell University and Blue Cross Blue Shield. The average age of NIH's extramural principal investigators is 51, and their intramural counterparts average a venerable 54.

3, 7, 31, and Counting …

Prime numbers—numbers divisible by only themselves and 1—are the "atoms" of mathematics. Mersenne primes, of the form $2^P - 1$ (where P is itself a prime), are special because of the computational and theoretical challenges they pose. It's not known, for example, whether the list of Mersenne primes goes on forever or if it stops at some largest exponent.

At the moment, Mersenne primes are piling up slightly faster than the overall trend would predict. The far right point on this graph was found 23 August by computers at the University of California, Los Angeles, running algorithms from the Great Internet Mersenne Prime Search (GIMPS). It corresponds to expo-

MERSENNE PRIMES

nent $P = 43,112,609$ and weighs in at 12,978,189 digits. That makes it the winner in a contest by the Electronic Frontier Foundation (EFF), which put up $100,000 for the first known prime with 10 million digits or more. The previous record, found 2 years ago, fell just short at 9,808,358 digits.

Computers in Langenfeld, Germany, came in 2 weeks too late for the big prize, finding the next Mersenne prime, with exponent $P = 37,156,667$, on 6 September.

The computations to verify primality get exponentially harder as the exponent increases, but GIMPS is relying on exponential growth of the Internet to keep pace. One of the next few Mersenne primes may reel in the next EFF prize: $150,000 for a 100-million-digit prime.

Roman Holiday

The Roman city of Herculaneum perished along with Pompeii when the 79 C.E. eruption of Vesuvius buried it under 20 meters of volcanic ash. Now, visitors can immerse themselves in a virtual-reality Herculaneum with the opening of the Museo Archeologico Virtuale (MAV), the first Italian virtual archaeological museum, on the site of the old city.

Created by engineer Gaetano Capasso, MAV includes more than 70 multimedia installations, including 3D screens and holographic projections. Its heart is the "cave," a room that features simulations of entire environments such as a room, theater, or garden. There is also a reconstruction of a Roman bath with real steam, where visitors can breath in the warmth and scent of ancient perfumes and gaze at holographs of women bathing. They can also don electronic gloves that let them dip fingers in a virtual fountain and feel the surfaces of virtual mosaics and fabrics.

The museum is the last word in virtual reality, says Derrick de Kerckhove, former director of the McLuhan Program in Culture and Technology at the University of Toronto in Canada. "No archaeological reconstruction has gotten deeper technological and artistic attention than this," he says.

HUMANE LENS. A former New York City fashion photographer hopes his camera can help end deadly attacks against albinos taking place in Africa.

Rick Guidotti, 51, has spent a decade photographing children and adults around the world with the rare genetic disease, which leaves eyes, skin, and hair without normal pigment and is a source of stigma. This passion left him determined to try to stem the violence in Tanzania, where at least 19 people with the disease have been killed in the last year as part of a brutal trade in albino body parts.

Campaigns

"I don't think we're going to stop everything with this one trip," says Guidotti, who leaves on 19 October on a campaign sponsored by Under the Same Sun, an albinism advocacy group in Canada. But, he says, he hopes to "demystify [albinism] and take away its oneness" by meeting with officials and sharing his photographs. Pictures of children with albinism and other genetic diseases can be seen on his Web site www.positiveexposure.org.

PIONEERS

HIGHER. Talk about an upward trajectory. Elon Musk lasted only days in his graduate physics program at Stanford University after he enrolled in 1995. But he's been no slouch since, having created and sold two major Internet companies and invested millions in solar energy and electric cars. Now 37, the South African–born Musk achieved a dream last month by sending the first private rocket into orbit.

SpaceX, which he founded in 2002, launched its Falcon 1 rocket from an island in the central Pacific. The 28 September flight followed three failed attempts. The rocket can carry up to 1100 kg, with a larger version rated to 12,500 kg. SpaceX plans to fly Malaysian and Canadian science experiments into orbit and has contracts with NASA and the U.S. military to fly payloads into space.

POLITICS

FRESH AIR. Barbara Hogan was serenaded recently by AIDS advocates in Cape Town, South Africa, celebrating her appointment as health minister by incoming president Kgalema Motlanthe.

The impromptu choir also hailed the departure of her predecessor, the controversial Manto Tshabalala-Msimang, who earned the ridicule and wrath of the HIV/AIDS community for downplaying the effectiveness of

antiretroviral drugs and promoting alternative treatments such as lemon juice and beetroot. "We're ecstatic about her ousting," says Nathan Geffen, a member of the Treatment Action Campaign, who was a part of the choir.

Hogan has been a member of the South African Parliament since 1994 and has worked on economic issues. A presidential election is scheduled to take place in 7 months, which leaves her little time to repair what Geffen sees as the "disaster" caused by Tshabalala-Msimang. "But Hogan can lay the groundwork for change to happen," he says, suggesting that she can start to heal the rift between civil society and the health department. And she could be kept on if Motlanthe retains power. "Many of us are hoping that things will play out quite well."

THEY SAID IT

"The funding provided by the United States for the construction of the Large Hadron Collider (LHC) does not constitute a 'major Federal action' as defined by the National Environmental Policy Act."

—Judge Helen Gillmor of the U.S. District Court in Hawaii dismissing a lawsuit by Walter Wagner and Luis Sancho, who believe that smashing beams at high energies in the newly commissioned accelerator at CERN in Switzerland might endanger the planet. The 26 September ruling did not address the question of whether the LHC poses a threat but noted that U.S. courts do not have jurisdiction over the facility.

Awards >>

TRIESTE PRIZE. A Brazilian astrophysicist and an Indian aerospace engineer are co-winners of the $100,000 Trieste Science Prize from the Academy of Sciences for the Developing World.

Beatriz Barbuy, a professor at the Institute of Astronomy, Geophysics and Atmospheric Sciences at the University of São Paulo, Brazil, studies the chemical composition of stars and has shown that metal-poor stars in the halo of the Milky Way have an overabundance of oxygen. The work has enabled a better understanding of how galaxies form.

Roddam Narasimha, chair of engineering mechanics at the Jawaharlal Nehru Centre for Advanced Scientific Research in Bangalore, India, and a professor at the University of Hyderabad, analyzes the role of turbulence in applications related to aerospace and atmospheric sciences. His research has helped improve aircraft design and monsoon predictions.

Got a tip for this page? E-mail people@aaas.org

NEWS>>
THIS WEEK

Zebrafish
development
in detail
176

Race: a hidden
voting bias?
180

NOBEL PRIZE IN PHYSIOLOGY OR MEDICINE

HIV, HPV Researchers Honored, But One Scientist Is Left Out

Nobel Prize stories often recount the early-morning phone call that startles a winner from slumber, followed by disbelief, and finally elation. When Robert Gallo's phone rang early on the morning of 6 October, he felt everything but elation. As a reporter explained to him, the Nobel Prize in physiology or medicine had been awarded to Luc Montagnier and Françoise Barré-Sinoussi of the Pasteur Institute in Paris for their discovery of the virus that causes AIDS. A third prize went to German

German Cancer Research Centre in Heidelberg, for his pivotal work on HPV. Unfortunately, say some of his colleagues, zur Hausen's well-deserved prize has been overshadowed by the controversy surrounding the choice of his fellow laureates.

Gallo, who during the past decade made peace with the Pasteur researchers after a prolonged battle over credit, reacted graciously to the news, noting that all three winners deserved the prize. But Gallo, head of the

in that," says Montagnier. "I'm very sorry for Robert Gallo."

No one questions the award to Montagnier and Barré-Sinoussi. The assembly credits them with isolating the human immunodeficiency virus (HIV) from a French patient with swollen lymph nodes in 1983. The researchers also detected activity of an enzyme called reverse transcriptase, proof that the infectious agent belonged to a group called retroviruses, which insert their own DNA into the genome of the hosts.

But when Montagnier, Barré-Sinoussi, and co-workers first described HIV in the 20 May 1983 issue of *Science*, they concluded that the "role of the virus in the etiology of AIDS remains to be determined." The first convincing evidence that HIV was the cause came 1 year later from Gallo and co-workers, who published four papers in *Science* that persuasively tied similar viruses they had found to the disease.

Gallo's lab discovered the first human retrovirus in the 1970s and then identified a biochemical that played a critical role in developing the technology to grow HIV in culture dishes. Gallo and colleagues subsequently made fundamental discoveries about the genes of the virus and how it enters cells. "I am sad that Bob was left out," says Nobel laureate David Baltimore, who now studies HIV/AIDS at the California Institute of Technology in Pasadena. "His work was critical to pinning down HIV as the cause of AIDS."

The rift between Gallo and Montagnier, fundamentally about apportioning credit, exploded after the 1985 issuing of patents for the HIV blood test. In 1987, U.S. President Ronald Reagan and French Prime Minister Jacques Chirac calmed the legal waters by proclaiming the two researchers "co-discoverers" of the virus and agreeing to split the patent royalties between the two countries. Gallo and Montagnier accepted that compromise and buried the hatchet, even co-authoring a chronology of AIDS research that ran in *Nature* that year. The controversy reignited in 1990, when the *Chicago Tribune* published a long investigative story by John Crewdson, who questioned whether Gallo's lab had stolen the virus from Montagnier's group. The U.S. National Institutes of Health, Gallo's employer at the time, investigated, and the ▶
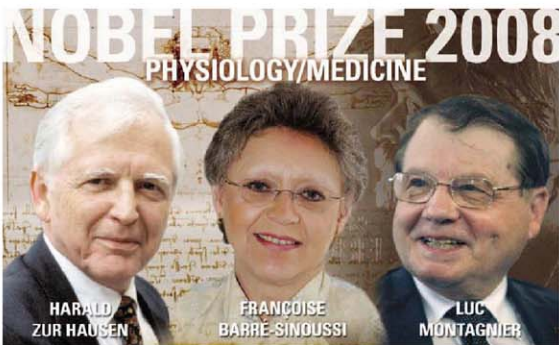
Virus hunters. Germany's zur Hausen discovered the link between human papillomaviruses and cervical cancer; France's Montagnier and Barré-Sinoussi first described the virus now known as HIV.

virologist Harald zur Hausen for finding that human papillomaviruses (HPVs) can cause cervical cancer. Given that Nobel Prize rules allow only three researchers to share an award each year, this meant that Gallo, whose lab has done pathbreaking research about AIDS since the epidemic surfaced, would join the long list of accomplished researchers whose work has been snubbed by the Nobel Assembly.

The assembly, 50 professors at the Karolinska Institute in Stockholm, Sweden, gave Montagnier and Barré-Sinoussi half of the $1.4 million award. The other half goes to zur Hausen, who is a professor emeritus at the

Institute of Human Virology at the University of Maryland School of Medicine in Baltimore, Maryland, acknowledged that he was "disappointed" to be left out. "Yes, I'm a little down about it but not terribly," Gallo told *Science*. "The only thing I worry about is that it may give people the notion that I might have done something wrong."

Montagnier, who now heads the World Foundation for AIDS Research and Prevention in Paris, told *Science* that he, too, was "surprised" that Gallo was skipped. "It was important to prove that HIV was the cause of AIDS, and Gallo had a very important role

FOCUS

Climate science
and the next
president
182

Celebrating Japan's
greatest mathematician
185

U.S. Congress became involved. But Gallo and his team were ultimately cleared of wrongdoing.

After their rapprochement in the late 1990s, Gallo and Montagnier each wrote essays in *Science* in 2002 in which they concluded that both made important contributions to the discovery of the virus. "Over the past 20 years, the scientific and legal controversies between our team and Gallo's group have faded," Montagnier wrote. The essays were seen by some as a way to prepare the ground for a shared, controversy-free Nobel award. Yet, "the protagonists don't get to write history themselves," says Pasteur researcher Simon Wain-Hobson.

Jan Andersson, an infectious-diseases specialist and a member of the 2008 Nobel Committee (the "working body" of the assembly), says that in making this decision, the panel is following Alfred Nobel's will by honoring the first discovery and not everything that followed. One reason that the HIV Nobel comes so late, Andersson says, is that it took a long time to tease out each scientist's role and forge a consensus in the assembly. That the rival labs resolved the issues between them did not matter, adds the committee's secretary, Hans Jörnvall. "When it comes to deciding who made Nobel Prize–worthy discoveries, I think we are the experts, not a set of lawyers," he said Monday during a press conference in Stockholm.

Gallo, once known for his hot temper and fierce competitive streak, told *Science* that he has mellowed a lot, making it easier for him to accept the Nobel Assembly's decision. "Twenty-five years ago, I'd be stuttering and saying, 'What the hell is going on?' As long as everything is not taken away from my legacy, that's fundamentally what matters to me."

Unlike the discovery of HIV, in which several competing labs fingered HIV early

Four's a crowd. The Nobel Assembly surprised many by not sharing the award with pioneering AIDS researcher Robert Gallo.

on as the cause of AIDS, zur Hausen was alone when he asserted in the 1970s that HPV causes cervical cancer. At the time, many researchers suspected that the herpesvirus and other sexually transmissible viruses were responsible for cervical cancer; zur Hausen bet on HPV, which was known to cause warts. "Many people thought he was out of his mind," says Philip Davies, director-general of the European Cervical Cancer Association. His discovery laid the groundwork for the development of a cervical cancer vaccine that is now being widely introduced in the United States and Europe.

In 1983, zur Hausen demonstrated the presence of HPV-16 and -18—two of many different HPV types—in cervical cancer biopsies. The two viruses were subsequently found in about 70% of cervical cancer biopsies throughout the world. "He showed the importance of innovative thinking in science," says HPV researcher Joakim Dillner of Lund University in Sweden. John Schiller, one of several key researchers who built on the discovery to help develop an HPV vaccine, says it's "great" that zur Hausen and his team have received the recognition for their critical work. "It is gratifying to see this ultimate payoff for their insight and perseverance," says Schiller, who works at the U.S. National Cancer Institute.

HPV has since been implicated in cancer of the vulva, anus, penis, and mouth. Zur Hausen's findings could have a huge impact on public health. In many developing countries, cervical cancer is the leading cause of cancer deaths among women. The World Health Organization, among others, is now lobbying for the vaccine's introduction in these poor countries.

Zur Hausen says he was "completely surprised" when he heard the news on 6 October shortly before the press conference in Stockholm began. And, of course, he was elated, too.

–JON COHEN AND MARTIN ENSERINK

NOBEL PRIZE 2008
PHYSICS
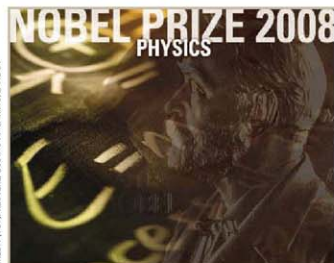
TRIO OF PARTICLE THEORISTS LAUDED

This year's prize honors the pioneering use of a key conceptual tool and an inspired guess. In the 1960s, Yoichiro Nambu, 87, of the University of Chicago in Illinois applied the notion of spontaneous symmetry breaking—the sort of thing that happens when the ions in a magnet align in some random direction—to particle physics. The concept now lies at the heart of the so-called standard model. In 1972, Makoto Kobayashi, 64, now at the High Energy Accelerator Research Organization in Tsukuba, and Toshihide Maskawa, 68, of Kyoto University, both in Japan, found that they could explain a slight asymmetry between matter and antimatter discovered 8 years earlier if there were six types of particles called quarks. Only three were known, but the others have since been found and the theory has been precisely confirmed. The award was announced as this issue of *Science* was going to press. For a more complete description, go to www.sciencenow.org.

DEVELOPMENTAL BIOLOGY

Lights! Camera! Action! Zebrafish Embryos Caught on Film

For understanding what really goes on as an embryo develops, it is hard to beat a good movie, and this week German researchers released a set of potential blockbusters. The unusual movies, described this week online in *Science* (www.sciencemag.org/cgi/content/abstract/1162493) and available on the Web, show all the movements and divisions of cells in a zebrafish embryo during its first day of development. By the end of the movies, the fish have a recognizable backbone and heart as well as the early stages of several other organ systems.

In previous years, *Caenorhabditis elegans* has starred in similar movies as fascinated developmental biologists watched the entire development of the 1000-cell roundworm. Other cinematic hits featured the formation of specific tissues and organs in fish, chicks, and mice. But until now, no one has managed to capture such detail for so long in an entire complex vertebrate.

The German filmmakers' technique promises to open new windows into the development of embryos and tissues in a wide assortment of species. "This is what we have always wanted to do—to follow everything in time and space," says Markus Affolter, a developmental biologist at the Biozentrum in Basel, Switzerland.

To make the movies, physicists Philipp Keller and Ernst Stelzer, with developmental biologists Annette Schmidt and Jochen Wittbrodt of the European Molecular Biology Laboratory (EMBL)

Starring role. A new set of movies reveals the precise cellular movements in live zebrafish embryos. These colorized "stills" show the embryo between 1.5 hours and 20 hours of development.

in Heidelberg, Germany, crafted a microscope that exposes living specimens to a laser-generated sheet of light. The micrometer-thin beam moves quickly enough through the specimen that the microscope can capture real-time images, sparing an embryo the bleaching and light damage produced by other microscopy techniques. Because of that bleaching, researchers were previously limited to following only a subset of cells for a few hours. The new design uses a digital scanner instead of a cylindrical lens to generate the light sheet. That produces clearer images of larger specimens and allows rapid scanning of an entire organism, Keller says.

The microscope allows observations of live specimens as big as 5 millimeters or more, and the researchers have already had success filming later stages of the zebrafish embryo, as well as embryos of chick, mouse, and *Xenopus* frogs. The frog embryo "is extremely challenging because it isn't very transparent," but the method does work, Keller says. The researchers hope eventually to collect a menagerie of such embryo movies, enabling cross-species comparisons that so far have been difficult if not impossible.

The zebrafish starring in the films received an injection of mRNA just after fertilization; this produced a fluorescent marker that in turn tagged chromatin. These tags enabled the researchers to use image-processing software to trace individual cells as they divide and migrate to form the first structures of the embryo. The movies can also be run backward, enabling researchers to trace cells in various organs back to their origins. In one example, the

"The revolution of live imaging has barely started. We are going to have a lot of fun."
—MARKUS AFFOLTER, BIOZENTRUM

German team identifies the cells that will eventually form the fish's retina.

The films have already helped answer a long-standing question about the formation of the zebrafish mesendoderm, the cell layer that eventually gives rise to most of the body's internal organs. About 5.5 hours after fertilization, some one-third of the embryo's cells—about 1550—tuck inside the ball-like embryo during a 2-hour period to form this cell layer. Scientists have debated whether the mesendoderm-destined cells travel in a synchronous wave or more independently. Wittbrodt and colleagues found that both are correct—with one kind of movement on the dorsal side, and the other on the ventral side of the embryo.

Recording the development of a mutant zebrafish strain called *one-eyed pinhead*, the researchers saw that only about 60 of the mesendoderm-destined cells made the proper migration, explaining the mutant's well-known problems in this stage of development. Such precise observations are "breathtaking," says developmental geneticist Lilianna Solnica-Krezel of Vanderbilt University in Nashville, Tennessee.

In their paper, the researchers include a blueprint for the microscope and a cost estimate for the parts: about €80,000 ($100,000). "It would be on my Christmas list," says developmental biologist Richard Adams of the University of Cambridge in the United Kingdom.

Even if other researchers can't afford their own microscope, they will be able to analyze the zebrafish movies and the data behind them by accessing a free Web site.* Much like physicists exploit the particle-track data generated by CERN and other particle smashers, scientists from around the world will download the zebrafish data and glean their own insights from it, predicts developmental biologist Mark Cooper of the University of Washington, Seattle.

"The revolution of live imaging has barely started," Affolter predicts. "We are going to have a lot of fun."

–GRETCHEN VOGEL

*www.embl-heidelberg.de/digitalembryo/
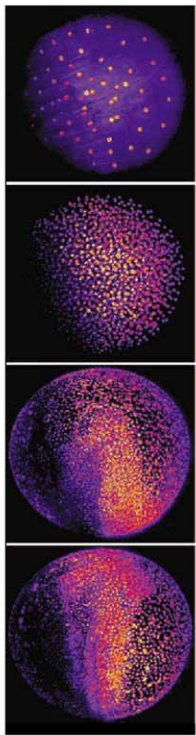
ENDANGERED SPECIES ACT

Pacific Northwest Sea Bird May Lose 'Threatened' Status

The U.S. Fish and Wildlife Service (FWS) agreed last week to review the protected status of the marbled murrelet, a sea bird that nests in the coastal forests of the Pacific Northwest. The move comes in response to a timber industry–led petition claiming that the bird does not meet key provisions of the Endangered Species Act of 1973 (ESA). But opponents cry foul, arguing that the industry petition hinges on a flawed analysis. What's more, other, more recent studies show that the population is in rapid decline, some scientists say, and any change to its status could be disastrous.

The fate of the small sea bird could have big implications for old-growth forests. To protect the marbled murrelet and also the spotted owl, whose habitat overlaps, up to 89,000 hectares of old-growth forests in the Pacific Northwest are off-limits to logging. Delisting the marbled murrelet could remove that barrier.

The controversy hinges on whether the birds on the U.S. side of the border are a "distinct" segment of the broader population of marbled murrelets that stretches across the northwestern part of the continental United States and into British Columbia. Yes, said FWS when it first listed the populations in California, Oregon, and Washington as threatened on 1 October 1992. The ESA mandates that FWS conduct updates on listed species every 5 years. The final version of a 2004 FWS update came to the opposite conclusion about the murrelets' distinctness, saying that the populations are not genetically distinct, and that all were protected by regulations on both sides of the border. Under ESA, several criteria can be used to judge distinctness: genetics, ecological role, or protected status.

Noah Greenwald, science director of the Center for Biological Diversity in Portland, Oregon, alleges political interference, citing a 28 February 2007 memo between FWS's Portland and Washington, D.C., offices acknowledging that Julie MacDonald, former deputy assistant secretary of the interior, had intervened in the 2004 report. "The Service recommended that the listing status remain the same, and supported listing as a distinct population segment," the memo reads. MacDonald and her office later overruled that recommendation. MacDonald has since resigned.

To Ann Forest Burns, vice president of the American Forest Resource Council

Unique? Continued protection of the U.S marbled murrelet hinges on whether it is "distinct" from its Canadian counterparts.

(AFRC) office in Portland, the 2004 study means murrelets in California, Oregon, and Washington should be delisted. In its petition, AFRC is now calling on the service to act on its previous findings. "We're trying to get some logic and transparency involved here," Burns says.

Not so fast, says Steven Beissinger, a conservation biologist at the University of California, Berkeley, who surveyed the marbled murrelet from 1995 to 2005. He says that even with the existing protections, the birds are suffering from "a bit of a triple whammy": lost habitat from logging, lost food resources, and failed nests due to predation. Central California populations have declined by as much as 70% in the past 2 years.

Meanwhile, a U.S. Geological Survey report released in 2007 concluded that certain segments of the murrelet populations in central California and Alaska are genetically distinct from the larger group, and populations across the range are plummeting.

FWS has 1 year to reach its much-anticipated decision.   **–RACHEL ZELKOWITZ**

CREDIT: MARK MOFFETT/MINDEN PICTURES/GETTY IMAGES

Change at the Top

The United Kingdom has brought the fight against global warming to the highest level of government. In a cabinet reshuffle last week, Prime Minister Gordon Brown chipped off parts of two existing departments to create a new Ministry for Energy and Climate Change. "I'm very much in favor," says David King, former U.K. government science adviser now at the University of Oxford.

The U.K. has struggled to pin down a domestic policy that both reduces carbon emissions and ensures energy security at a low cost. But with the business ministry in charge of energy and the environment department handling climate change, policy was often pulled in different directions. King says the new ministry should improve cohesion, but he hopes the new bureaucracy won't slow the progress of legislation. Ed Miliband, a former minister in the Cabinet Office and younger brother of the U.K.'s foreign secretary, will lead the new ministry.   **–DANIEL CLERY**

Show Me the Rubles

Last week, Russia's science ministry debuted a new scheme to evaluate 2500 Russian research institutes and guide funding decisions, but some scientists say it's a mistake. Following a new policy of making science effectively contribute to Russia's economic development, the scheme includes criteria such as the number of articles published by the staff, their citation impact factor, the number of international contracts, and even the gender proportion of the staff. Alexander Naumov, co–deputy director of the ministry department for scientific, technical, and innovation policy, says the assessments "will be taken into account by federal executive bodies [and] state science academies" when setting budget and planning priorities.

Russia's research funding system may need restructuring, say critics, but the new method is the wrong answer. "I tried to evaluate, according to this method, the organizations that I know," says Andrey Finkelstein, director of the Institute of Applied Astronomy of the Russian Academy of Sciences in Moscow. "One of the biggest astronomical organizations in the country appeared to be in the last place, while a very insignificant one turned out to be the most advanced one." After public comment, the ministry plans to amend and adopt the new scheme by November.

–ANDREY ALLAKHVERDOV AND VLADIMIR POKROVSKY

Meeting of Research Leaders Spotlights African Development, Disaster Planning

MAPUTO, MOZAMBIQUE—Battered by cyclones, prone to major floods, and still recovering from a lengthy civil war, this southeastern African country might seem like an unlikely meeting spot for the world's learned societies. But the International Council for Science (ICSU), bypassing the usual tourist magnets, chose to hold its general assembly session on 20 to 24 October in Mozambique's scarred but rebuilding capital, Maputo. ICSU did so in part to advance its strategy of expanding activities and boosting science in Africa, says its executive director, Swedish microbiologist Thomas Rosswall. "We are reaching out to the scientific communities in the developing world," he says, noting that ICSU has opened regional offices in Africa, Latin America, and Asia, and is planning a fourth in the Arab world.

The goals of the general assembly meeting, the first held in sub-Saharan Africa by the 77-year-old ICSU, are ambitious. Heading the list is a campaign to set priorities and help coordinate the work of interdisciplinary groups

researching disaster risks. The expected 250 or so attendees—representing 29 international scientific societies as well as national academies or agencies from 134 countries—will help launch the research. Although ICSU does not directly fund them, it helps plan and coordinate major international research projects. The newest such global collaboration, the hazards initiative, is akin to the International Polar Year of 2007, in which ICSU played a key role. At the Maputo meeting, the organization will also install French physicist Catherine Bréchignac, former director-general of CNRS, as its new president, replacing chemist Goverdhan Mehta of India.

Mozambique is an appropriate place for discussing disasters. It has been struck repeatedly by severe cyclones and floods. "African science needs ICSU," says computer scientist Patricio Sande, president of the Scientific Research Association of Mozambique and ICSU's local point man. In the years since its 16-year civil war, he says the country has made strides in rebuilding research, establish-

ing a science ministry 3 years ago.

Mozambique and other African countries "recognize the importance of science in dealing with their own development challenges," says Bréchignac. "It is important for developing countries to be involved in global scientific challenges."

Disaster planning is a good example, says climate scientist Gordon McBean of the University of Western Ontario in Canada, who will present the findings of ICSU's disaster risk group. He says "the poorest people are often the most affected by natural disasters," which he expects will become more frequent and severe with climate changes. Indeed, climate change is a major topic in Mozambique, whose long coastline would be vulnerable to rising sea levels. Deliang Chen, a climate scientist at the University of Göteborg in Sweden who will become ICSU's next executive director in February, says ICSU "has a responsibility in mobilizing the global scientific community" to do badly needed regional studies in Africa.

ICSU's focus on Africa has started to pay off, Rosswall says: The number of African countries with national members in ICSU has increased from 16 to 25 in the past few years. Microbiologist Khotso Mokhele, former head of South Africa's National Research Foundation, says ICSU had "a gaping hole in devel-

Comprehensive Conservation Database Details Threats to Mammals

For years, biologists and conservationists have expressed alarm at the plight of marine mammals and many large land mammals. But there's been no comprehensive look at how the world's mammals are faring—until now. Earlier this week, at its quadrennial congress, in Barcelona, Spain, the International Union for Conservation of Nature (IUCN) released its long-awaited database detailing the status of all mammals known to humankind since the year 1500. On page 225 of this issue of *Science*, the team that assembled the database analyzes the findings. The news is bleak, particularly for the oceangoers.

Overall, the analysis estimates that at least one-quarter of the world's 5487 known mammalian species are threatened with extinction—but an

estimated 36% of the marine mammals are at risk. Although all mammals face habitat loss and hunting pressures, for marine species, accidental death—being struck by ships or caught in fishing gear—poses the most widespread threat, IUCN reports.

The database, part of the IUCN Red List

Mammals at risk. Locally extinct in some places, harp seals face future threats from melting sea ice.

of Threatened Species, updates and expands a 1996 survey. The effort took 5 years and involved more than 1700 researchers from 130 countries. They combed the literature and pooled their unpublished knowledge of ecology, taxonomy, distribution, population trends, threats, and conservation efforts. The species were then classified according to their extinction risk. This new database has expanded coverage of marine mammals and of other groups, such as rodents, not generally thought of as being in dire straits. "We wanted to make this one-stop shopping for scientists and policymakers," says IUCN and Conservation International mammalogist Jan Schipper, who coordinated the project.

The good news is that well-funded conservation programs are succeeding. One example cited by Schipper is the recovery of the black-footed ferret in the United States, where it had disappeared from the wild and has been reintroduced from captive breeding programs (*Science*, 12 May 2000, p. 985).

Maputo rising. Mozambique's capital, with a boost from computer scientist Patricio Sande (*inset*), is hosting a gathering of international scientific unions.

oping countries" when he first got involved with the organization in 1993. He says the decision to hold the general assembly in Maputo is a "very significant" demonstration of ICSU's new commitment.

Mohamed Hassan, a Sudanese mathematician who is president of the African Academy of Sciences and executive director of the Academy of Sciences for the Developing World, says, "You are now seeing in Africa the emergence of a new breed of leaders who see the importance of science and technology to their futures." Although it took ICSU a long time to improve its developing world contacts,

Hassan says, he thinks its regional offices have been a valuable addition.

"ICSU punches above its weight," says Mokhele, noting that ICSU played a role in the formation of the Intergovernmental Panel on Climate Change and the coordination of the International Polar Year. "Even though it is smaller than UNESCO [United Nations Educational, Scientific and Cultural Organization] or the other huge nongovernmental organizations, it is capable of doing things that the others can't do."

–ROBERT KOENIG

Also encouraging is the diversity of mammals, which is richer than previously thought. Taxonomists have described 349 newly discovered mammals since 1992, including an elephant shrew from Tanzania early this year. All told, there were 700 mammals in this survey that were not covered in 1996.

The bad news is that 188 species are critically endangered, and 29 of them, such as a freshwater dolphin from China called a baiji (*Science*, 22 December 2006, p. 1860), may already be extinct. Since 1996, the risk categories and mammal classification schemes have changed, making it difficult to directly compare 1996 data with the new information, says Schipper.

Large land mammals, particularly hoofed animals and primates in South and Southeast Asia, are the worst off, Schipper's team reports. Marine mammals are most at risk in the North Pacific, the North Atlantic, and Southeast Asia.

All told, extinction looms large for 1139 out of the 4651 species for which there are good data. The other 836 species were so poorly studied that it was impossible to tell what their conservation needs might be. About half the species are declining in numbers, including one in five of those not at risk of extinction right now.

The new database "is the most valuable effort to date to summarize the state of conservation and threats to the world's mammal populations," says mammalogist Don Wilson of the Smithsonian National Museum of Natural History in Washington, D.C. "By detailing threats at the species level, it will now be possible for management agencies in every country in the world to prioritize their efforts to try to mitigate these threats."

There will likely be a payoff for basic research too, says mammalogist Gerardo Ceballos of the National Autonomous University of Mexico, who in 2005 compiled a similar database encompassing about 4500 mammals. These databases can reveal large-scale patterns of species distributions, Ceballos says, and their relationship to climate change and extinctions.

–ELIZABETH PENNISI

Reviewing Peer Review

In a 2-year experiment it hopes to extend to other disciplines, the Italian Ministry of Research and Health last week announced the creation of a Science Committee that will implement a new peer-reviewed funding system for biomedical research. The move comes in response to Italian scientists' demands for more transparent funding decisions uninfluenced by politics or favoritism. The panel will choose independent referees to evaluate grant and project proposals, and those referees' judgments will then be considered annually by an international expert panel. The U.S. National Institutes of Health will help set up the program. Gilberto Corbellini, a professor of the history of medicine at the University of Rome, La Sapienza, says the reforms are needed, but "a law to prevent the conflict of interests" is required to prevent politicians from nominating members of the Science Committee.

–LAURA MARGOTTINI

Target Acquired: Lung Cancer

The U.S. Army is figuring out how to spend $20 million that Congress has given the Department of Defense (DOD) for peer-reviewed research on lung cancer. In a 2009 budget bill passed in late September, lawmakers added lung cancer to a list of cancers that DOD must study. The Lung Cancer Alliance (LCA), which helped push for the earmark, hopes the bill's emphasis on "early curable lung cancer" will support a search for biomarkers and standards for computed tomography (CT) lung scans to detect small tumors. These tests increase costs for follow-up procedures; it's not clear that the scans reduce mortality (*Science*, 2 May, p. 600). But LCA President Laurie Fenton Ambrose says that many patients are already getting lung CT scans and that standards are urgently needed.

–JOCELYN KAISER

Report: U.K. Physics Physically Fit

Despite concerns that gripped British physicists last year, the U.K. physics enterprise is healthy, says a report published by the country's research councils last week. The review, by chemical engineer Bill Wakeham of the University of Southampton, found that U.K. physics had high international standing with a citation impact second in the world. But it recommended strengthening the pipeline of researchers through school and university, training more physicists as teachers, and coaxing more women and minorities into physics.

–DANIEL CLERY

SOCIAL SCIENCE

Do Voter Surveys Underestimate The Impact of Racial Bias?

As a grueling, multiyear presidential campaign comes to an end, a touchy question hangs in the air: Have voter polls been skewed by an inability to detect racial bias? Some social scientists think that many voters who say they support Democratic candidate Senator Barack Obama in fact may be uncomfortable with the prospect of an African-American president—or that polls have failed to reach Democratic voters most likely to harbor such prejudice. On election day, the theory goes, some fraction of those people may vote for Republican Senator John McCain—or not vote at all.

A few social scientists are trying to quantify this hidden racial bias. But in presidential politics, they're traversing new terrain. "Since polling began, there's never been this situation," with a white presidential nominee facing off against a black one, says sociologist Maria Krysan of the University of Illinois, Chicago. Although pollsters assume that some voters reject Obama because of his race, they don't know whether the polls are capturing all of that bias. One voter model that does attempt to factor in hidden racial bias, being published this month, suggests that it may drag down the numbers for Obama by 6%.

Science and the 2008 Campaign

Concerns about this hidden factor, commonly called the "Bradley effect," date back to 1982 when Tom Bradley, an African American, lost the race for governor of California despite being ahead in polls prior to the election. A similar pattern showed up in 1989, when African-American Douglas Wilder had a 15% lead in the Virginia gubernatorial race but scraped by with a half-percent win in the election.

Evidence for a Bradley effect is decidedly mixed in the presidential campaign this year. Many African Americans have won elections for mayoral, state, and congressional offices since the Bradley days. Daniel Hopkins, a political scientist at Harvard University, notes that "Americans' racial attitudes have shifted markedly" in this time. Still, Hopkins was intrigued by the Democratic New Hampshire primary in January, when Senator Hillary Clinton pulled off a decisive win even though Obama had surged ahead in the polls the week before.

Turning to past elections, Hopkins examined every gubernatorial and Senate race from 1989 to 2006 that included either an African-American or a female candidate, on the rationale that a woman could also experience voter prejudice. Of the 133 elections he analyzed, just 18 included an African-American. Hopkins found no Bradley effect for blacks after 1996, and none for women at all. Before that time, African-Americans appeared to lose 2% or 3% of support stated in polls. Hopkins presented the work in July. "As best I can tell, unless race really became a central issue in this political campaign, I don't expect a Bradley effect," he says.

Andrew Kohut, president of the Pew Research Center in Washington, D.C., which performs public opinion polls, rejects the argument that voters are dishonest when asked whether they'll support Obama. That's partly because asking someone whom they plan to vote for is different than "if you ask them about race head-on," he says. But Kohut does worry that polls may still be skewed, because studies have shown that what pollsters refer to as "reluctant respondents"—people less likely to participate in a poll—tend to have less favorable attitudes toward racial minorities, and their intent may be missed in current polls.

"We are in an unknown world; we don't know what's going to happen with a black candidate," says Michael Lewis-Beck, a political scientist at the University of Iowa in Iowa City. Like others in a small cohort of election predictors, he combines voter polling responses, economic data, and public perceptions of the White House incumbent to forecast presidential elections every 4 years. Of nine groups publishing their models this month in *PS: Political Science & Politics*, Lewis-Beck is the only one who has adjusted his model to take race into account. His fine-tuning is based on voting patterns in the primaries and on how honest individuals are in saying they will support an African-American presidential candidate.

His prediction? Obama is likely to corral 50.1% of the popular vote but lose the election in the electoral college. Running his model, without taking race into account, Obama would capture 56% of the vote, says Lewis-Beck, a figure that's higher than most of his colleagues give. (Six of the nine models predict that Obama will win with at least 52% of the vote.) "It's a feel-good thing; people don't want to admit that race is a factor," he says, noting that on average he's been off by about 1.5% in his predictions in the past. Those who disagree that race will have much effect on voting patterns, he says, "live in ga-ga land."

Few analysts have tried to quantify the problem of hidden bias, precisely because it is so hard to detect. Most polling is done by phone and involves a one-on-one interaction. "It's not a positive attribute in general to express racial prejudices to a virtual ▶

Obscured. Attitudes about race that aren't captured in surveys during the presidential election campaign may come to the fore when people cast their ballots in the privacy of the voting booth, experts say.

stranger," says Krysan. But bias has been detected. Krysan recalls a study she did back in 1994, asking white residents of the Detroit metropolitan area whether they would vote for an African-American candidate: 11% said in a face-to-face interview that they would not, and 22% said no in a mail survey, which they filled out privately.

A pair of graduate students tried to measure such racial attitudes in a June 2007 survey of 1500 people, whom they divided randomly into three groups. Brian McCabe and Jennifer Heerwig, sociology students at New York University in New York City, used statistical tools to unmask covert attitudes. They asked their control group of 500 people point blank whether they would vote for an African-American president, and 84% said yes—a

lower percentage than other polls. McCabe and Heerwig attribute the difference to greater privacy: Their survey was taken over the Internet. They asked the second group three questions unrelated to race, such as whether presidential campaigns are too costly. The third group was asked those same three questions, plus a fourth about whether they'd vote for an African-American candidate. McCabe and Heerwig knew only how many statements the respondents in those two groups agreed with, not which ones. "If you have a socially undesirable answer, you can hide it," says McCabe. With some statistical analysis, they determined that assuming the three groups were similar, 14% of their first cohort lied, and in truth just 70% would support an African-American. Democrats were

much more likely to be dishonest than Republicans, proffering up the socially desirable answer, as were those with less education.

However, it's not clear "how well a question about a generic black presidential candidate translates to Barack Obama in particular," says Heerwig. The two have presented the work at meetings but not yet published it.

Some analysts argue that racial bias is clearly visible, however. "I think it's already shown up in preference polls," says James Campbell, a political scientist at the University of Buffalo in New York state and the only one of the forecasters to predict that McCain will capture the popular vote. "This should be a very strongly Democratic year," he says, but McCain still has a shot at the White House. **–JENNIFER COUZIN**

Tax Credit Extension Is Silver Lining for Science

Last week's $700 billion rescue package for U.S. financial institutions also lends a hand to high-tech companies by extending and expanding a tax credit for investing in research.

The research credit was part of a cluster of tax measures grafted onto the rescue plan at the last minute to entice U.S. House of Representatives members who had rejected the bailout 4 days earlier. It provided some consolation for science lobbyists disappointed that Congress had wiped out proposed increases for several research agencies when it passed a 6-month freeze on federal spending before recessing for the 4 November election (ScienceNOW, 29 September, sciencenow. sciencemag.org/cgi/content/full/2008/929/1). Although the U.S. Senate will return next month for a brief lame-duck session, the House currently has no plans to do likewise.

A perennial legislative priority for businesses, the research credit was first adopted in 1981 and has been extended a dozen times, often for a single year, and periodically revised. The original law provided a 20% tax credit for additional research spending using a formula based on research as a share of overall sales from an earlier period. However, most companies now use a "simplified" credit adopted in 2006 that rewards them for increasing their research budgets.

Last week's legislation raises the percentage on the simplified credit from 12% to 14%; it also makes it retroactive to the date the old provision expired, 31 December 2007, and extends it through 2009. Although businesses have lobbied hard for a permanent, 20% credit, they are grateful for what Congress has done.

"At least it's not a gap, and it sends a loud signal that this is important. But it's not ideal," says Monica McGuire of the National Association of Manufacturers, part of a business coalition that has pushed the measure (investinamericasfuture.org).

The political uncertainty can discourage U.S. companies from investing in new research facilities at home, notes Patrick Wilson of the Semiconductor Industry Association. "If you want to encourage a change in behavior, you really need to make it permanent," he says. The idea already has widespread support: Both Barack Obama and John McCain say the credit should be permanent, as did presidents George W. Bush and Bill Clinton. But Congress has always balked at the idea because of the long-term cost, which under current legislative rules must be offset by new revenues. The credit will cost the U.S. Treasury an estimated $8.3 billion next year.

Meanwhile, science lobbyists are catching their breath before trying to figure out how to salvage the 2009 fiscal year, which began on 1 October. Their hopes were buoyed briefly late last month when the Senate took up a $56 billion stimulus package proposed by Senator Robert Byrd (D–WV), chair of the Senate Appropriations Committee, that would have given an additional $1.2 billion to the National Institutes of Health, $250 million more to NASA, and $150 million more to the Department of Energy's Office of Science. It also contained $925 million for the U.S. Coast Guard to build a polar icebreaker. (A 2006 report from the National Academies' National Research Council said two new vessels were vital for

national and economic security as well as for continued scientific exploration.) Although a majority of senators supported Byrd's bill, it failed to receive the 60 votes needed for passage under special rules. The same day, the House passed its own stimulus package that contained no money for research.

Unless the lame-duck session takes an unexpected turn, the next shot at boosting 2009 funding levels will come after the next president is sworn in and the new Congress takes up funding the government for the rest of the fiscal year. Depending on the economy, however, advocates may find that making the case for additional spending on science could prove even more daunting than it was in the fall. **–JEFFREY MERVIS**

Into the woods. Research on climate change impacts includes studying how forests will respond to rising temperatures.

strengthen Earth monitoring and efforts to link scientists and local officials.

To implement those changes, the next president will need to beef up and restructure the 18-year-old CCSP. A string of reports by experts say CCSP has been plagued by a stagnant budget, poor coordination between participating agencies, and a lack of White House leadership. The U.S. National Academies' National Research Council (NRC) concluded in 2007 that local and regional officials are receiving "inadequate" help in preparing for potentially catastrophic changes. Its report also pointed to the country's "relatively immature" understanding of how climate change may affect residents. "The health of the climate science [program] is not what it should be," says Representative Rush Holt (D–NJ), speaking on behalf of the Obama campaign.

Impacts Research Seen As Next Climate Frontier

Scientists hope the next U.S. president will devote more of the billion-dollar climate change research program to impacts

Marine ecologist Jane Lubchenco was among the first scientists to study how ecosystems off the California coast are being affected by climate change. Although that work has put her ahead of the curve, it's hurt her chances of obtaining funding from the $1.8 billion U.S. Climate Change Science Program (CCSP), the major federal effort in the field. "[Its] focus has definitely not been on understanding impacts," says Lubchenco, a professor at Oregon State University, Corvallis, and a former president of AAAS (which publishes *Science*). Instead, she's relied on grants from private foundations to support her examination of oxygen-depleted oceanic "dead zones."

Neither Republican John McCain nor Democrat Barack Obama has discussed climate change research on the campaign trail. But both presidential hopefuls have

Science and the **2008** **Campaign**

weighed in on the need to better understand the regional consequences of global warming—the kind of information Lubchenco is collecting. In May, at a town hall meeting in Portland, Oregon, McCain warned of "more forest fires" and "more heat waves afflicting our cities." In July, Obama told a Dayton, Ohio, audience that climate change could bring "devastating weather patterns, terrible storms, drought, and famine."

McCain is thinking about reorienting the climate research program toward what his aide, Floyd DesChamps, calls "urgent impacts." He says that the White House's "21 [CCSP] reports" are inferior to the "real National Assessment" that his boss would launch. Obama's campaign says he'll stress "short-term and long-term effects" on society and ecosystems. Both candidates have promised to

Missed opportunities

Created by Congress in 1990, CCSP coordinates climate change research across 13 federal agencies. Initially called the U.S. Global Change Research Program, it helped U.S. scientists lead a global effort that by 1999 had, in the words of former GCRP Director Richard Moss, "nailed the question of detection, 'Is climate changing?' and attribution, 'There is a human cause.' " The next step, Moss says, would be getting a better handle on the impacts of climate change and developing adaptation strategies.

In 2002, the Bush Administration renamed the program and laid out five overarching strategic goals, three addressing basic climate science and two focused on impacts and adaptation. But roughly 75% of the funding gets spent on the first three goals. "They said, 'Wait a minute, we're not there on the question of detection and attribution,' " says Moss, who ran CCSP until 2006. "There was far less of a shift in the program than [we had proposed]."

William Brennan, the current CCSP

director, says the lopsided emphasis within CCSP on characterizing global climate change over identifying impacts reflects "our state of [scientific] understanding." But the NRC study said the U.S. program lacks the investment in data or modeling capabilities to forecast how warming might create feedbacks, such as carbon released from warming soils or methane from melting tundra.

As did the Clinton Administration, the Bush team gave CCSP little power to set agency budgets or shift priorities. Each participating agency controls its own budget and must approve decisions taken by CCSP staff. "The structure discourages the CCSP coordination office from taking initiatives on anything that even a single agency, or a single White House official, opposes," says former CCSP staffer Nick Sundt. "It's a deficiency," acknowledges Brennan.

A report released in August by eight national climate and weather organizations says CCSP needs more budgetary control and recommends that its director "report directly to the President." Policy experts believe a stronger CCSP could persuade agencies to invest more in areas such as impacts. They also argue that the effort needs more money.

The bread-and-butter climate research budget has hovered at about $1.9 billion in constant dollars since 1994, although a recent downward trend has squeezed academic researchers and federal climate scientists alike (see graph). The totals include funds for satellite programs—more than $1 billion in some years. Infrastructure needs for labs run by the National Oceanic and Atmospheric Administration (NOAA), the National Weather Service, and NASA take up much of the rest, although the exact distribution is hard to follow. (A 2005 report by the Government Accountability Office, for example, is labeled *Federal Reports on Climate Change Funding Should Be Clearer and More Complete*.)

The declining amount of funds actually available for scientific research has meant fewer scientists tackling increasingly complex issues, including modeling ice sheets and measuring the effect of aerosols on local climates (*Science*, 22 August, p. 1032). A flat budget has also put a squeeze on hiring the next generation of scientists. As an example, the National Center for Atmospheric Research (NCAR) in Boulder, Colorado, receives 300 or so mostly "outstanding" applications each year for postdoctoral fellowship spots but can support only 40, says Jack Fellows of

the University Corporation for Atmospheric Research, which manages the center. "The system is so starved right now," adds Harvard University atmospheric chemist James Anderson.

Feeding it would require a budget of $4.5 billion by 2014, say climate research advocates, citing recommendations in reports by the U.S. Commission on Ocean Policy and NRC. But achieving that level won't be easy, admits Fellows, who helped write the August report and who briefed White House budget officials on its contents. "They kind of laughed at me," he recalls.

Having a voice

Researchers who run the nation's Earth-sensing network are feeling the pinch as much as anyone. Continuous measurements from space are crucial for monitoring climate impacts such as ice melting, sea-level rise, and changing pollution patterns, and such data underpin the whole

A DECLINING INVESTMENT
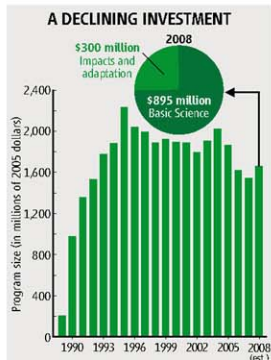
A basic difference. The CCSP's research budget is dominated by efforts to characterize and model global climate rather than understand its impacts.

climate research program. In addition, as support grows for implementing emissions cuts, "you have to monitor the planet closer than we're doing to see that [the cuts] are working," says NCAR Director Eric Barron.

The U.S. environmental sensor fleet includes roughly 30 orbiting satellites studded with more than 120 instruments, as well as land stations and ocean buoys. Most climate sensors fly on experimental NASA crafts designed to last roughly 5 years. To establish a more permanent

system, the government is building the $14 billion National Polar-orbiting Operational Environmental Satellite System (NPOESS). Managed jointly by NASA, the Pentagon, and NOAA, NPOESS includes 12 weather, climate, and space weather sensors on five bus-sized satellites, launched sequentially from 2010 to 2026.

It's an audacious vision, but the next Administration will inherit a 14-year-old program in trouble. In addition to soaring costs, repeated delays have increased the chances there won't be NPOESS satellites in orbit to continue crucial climate data records when NASA satellites fail. Scientists say an influential White House climate office might have made a difference. "We didn't play a role in that," says Brennan, who is also acting NOAA administrator. "I wasn't pleased."

Both presidential candidates have promised to revitalize Earth monitoring, but they'll also face the problem of processing NPOESS climate data on the ground. "The NPOESS program lacks essential features of any well-designed climate observing system," noted last year's NRC report.

Some critics of the U.S. climate research program say it needs more than an increase in funds and new hands on the helm. One radical government restructuring would merge NOAA and the U.S. Geological Survey to create an Earth Systems Science Agency. "No one agency has Earth observation as their number-one priority," says Mark Schaefer, a former official at the Interior Department, who proposed the idea in an article in *Science* earlier this year (*Science*, 4 July, p. 44) with former top brass from NOAA, NASA, and other agencies. A unified monitoring agency could manage huge programs like NPOESS more effectively, he claims.

The Bush Administration opposes such a massive reshuffling. "We need to focus on what needs to get done. I don't want to spend 2 years moving boxes around in the federal government," says Mary Glackin, NOAA's chief operating officer.

Scientists who have briefed the campaigns on how to reform the nation's current climate research say the candidates have gotten the message. The current CCSP staff is also preparing transition documents that recommend a shift toward impacts science. "The campaigns are talking about this," says Lubchenco, approvingly. "What I would like to see is that [talk] transformed into viable research programs."

–ELI KINTISCH

On golden-line pond. Yang Junxing's Ottelia (*inset*) and golden-line barbels (*pictured below*) are flourishing.

From Remarkable Rescue To Restoration of Lost Habitat

Southwest China's alpine lakes have lost many of their native species. Researchers may have found a way to reboot the ecosystem

KUNMING, CHINA—In a marshy pond the size of an Olympic pool in southwestern China, a lost world is flourishing. Golden-line barbels, a fish extirpated from its native habitat, dart among two other resurrected species: water plants with delicate white flowers called *Ottelia*, and freshwater mussels. "This is a healthy ecosystem," says Yang Junxing, deputy director of the Kunming Institute of Zoology (KIZ) of the Chinese Academy of Sciences. A dike separates the pond from a lapping water body that these species once called home: ailing and algae-ridden Lake Dianchi.

Pollution, overfishing, and alien species have devastated Dianchi, which at 310 square kilometers is the biggest lake in Yunnan Province and the sixth largest in China, and other nearby alpine lakes. "We are talking about a fish extinction comparable to Lake Victoria when Nile perch were added," says David Aldridge, a zoologist at the University of Cambridge, U.K. "It looks pretty massive." Twenty-one of Dianchi's 25 indigenous fish species are gone, having been supplanted by 31 alien species.

But thanks to a stroke of luck and dogged persistence, Yang's team may have hit on a way to bring Dianchi, just south of Kunming, Yunnan's capital, back into balance. Unlike many restoration efforts, Yang's may be economically viable. "Their plan for transforming Dianchi has caught the imagination of many of Yunnan's decision-makers," says World Bank biodiversity specialist Tony Whitten, who oversaw the team's work under a just-completed 5-year Global Environment Facility grant.

The situation at Dianchi was already grim when Yang was a graduate student in ichthyology in 1986. Golden-line barbels (*Sinocyclocheilus grahami*), 200 grams fully grown and once the lake's main commercial catch, had vanished. What doomed the wee fish is that people couldn't get enough of the delicacy. Then came the stroke of luck: In 1993, Yang found golden-line barbels in a "dragon pond"—a tiny spring at a Buddhist temple near Dianchi. "Thank Buddha!" Yang says. "Without the temple, there's nothing we could have done." Since then, four more fish snuffed out in the lake have turned up in dragon ponds.

But for more than a decade after their rediscovery, golden-line barbels would not breed in captivity. Then Yang's team at the Yunnan Endemic Species Breeding Center noticed that adults disappear in winter and reappear with hatchlings in spring, says KIZ ichthyologist Chen Xiao-Yong. He speculates that golden-line barbels breed in submerged caves during winter. "We don't know for sure. Its ecology is mysterious," Chen says. In 2007, "we learned how to simulate the cave," he says. They put cinder blocks in the breeding center's tanks for fish to hide in and fiddled with inputs such as fish chow and dissolved oxygen. Earlier this year, the team produced 80,000 golden-line fry. They are also raising two other endemic fish that are much easier to breed: white minnows (*Anabarilius grahami*), which by 2000 had been mostly wiped out of Yunnan's Fuxian Lake, and barb-

less carp (*Cyprinus pellegrini*). They've just started work on a fourth species, Dianchi bullhead (*Pseudobagrus medianalis*).
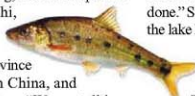
In the Xialiangwang experimental pond on Dianchi's eastern edge, Yang has reconstituted three flagship species: golden-line barbels, *Ottelia* (*Ottelia acuminata*), and mussels (*Anodonta woodiana*). A single *A. woodiana* can filter 50 liters of water a day, helping *Ottelia* flourish. The *Ottelia* thwart eutrophication by shading out algae, while golden-line barbels host *A. woodiana* larvae. "These species evolved to coexist," says Yang.

Dianchi itself is still a mess. Each year since the 1960s, Yunnan fishery departments have stocked Dianchi with four types of carp. One, the grass carp (*Ctenopharyngodon idella*), went on a rampage. In the 1970s and 1980s, these lawn mowers with gills chewed through Dianchi's *Ottelia* and other water plants, allowing blue-green algae to run riot. Other introduced fish devoured zooplankton. "Most fish species here now are introduced; it's very sad," says Chen. Eutrophication set in fast.

Dianchi's future could look much like Yang's experimental pond—if authorities consent to a restoration plan. Once the captive-bred golden-line barbel population surpasses 10 million, within a few years, Yang says, KIZ will petition the Yunnan Agriculture Bureau to approve the restoration and cease stocking Dianchi with carp. (Managers stopped purposefully adding grass carp in the 1990s, but grass carp fry are mixed in with fry of silver carp brought in from the lower Yangtze.) It will take a few more years, he says, to fish out the grass carp, which do not breed in the lake. Once they succeed, *Ottelia*, golden-line barbels, and mussels can be reintroduced en masse.

Yang argues that the restoration could even be profitable. The breeding center has begun selling limited quantities of golden-line barbels to wealthy customers for more than $100 per kilogram. If the fish were to flourish in Dianchi, fishers could rake in more money from golden-line barbels than from carp, the present favorite catch, Yang says, and locals could supplement their incomes by harvesting *Ottelia* stems for human consumption and its leaves for animal feed.

Testing the waters, Yang's group has reintroduced some 5000 golden-line barbels in cages in Dianchi. The fish are faring well, Yang says. But it will take millions more golden-line barbels—backed by an armada of *Ottelia* and plenty of mussels—to return Dianchi to its former glory.
　　　　　　　　　　　　　　–RICHARD STONE

CREDITS (TOP TO BOTTOM): R. STONE/SCIENCE; COURTESY OF YANG JUNXING

HISTORY OF SCIENCE

Samurai Mathematician Set Japan Ablaze With Brief, Bright Light

Isolated from the West, Seki Takakazu churned out some of the finest mathematical work of his time. Centuries later, scholars are finally giving him his due

TOKYO—Seki Takakazu, arguably Japan's greatest mathematician ever, labored in isolation at a time when Japan had cut itself off from the rest of the world. Yet his discoveries rivaled and often anticipated those of European mathematicians he never heard of. With the 300th anniversary of Seki's death approaching on 24 October, historians met here* to assess his work and discuss why the tradition he embodied—an indigenous Japanese mathematics known as *wasan*—languished after he was gone.

"There are many new discoveries" about Seki's life, says Shigeru Jochi, a specialist in early East Asian mathematics at National Kaohsiung First University of Science and Technology in Kaohsiung City, Taiwan. For example, Jochi says, Seki worked as a professional mathematician—contrary to previous thinking that mathematics in feudal Japan was a mere pastime. Hikosaburo Komatsu, a science historian at Tokyo University of Science, says Seki's most important discovery—a general theory of solving systems of equations—has been overlooked because it led to calculations "almost beyond human capabilities."

Seki was born into a samurai family around 1640 in Fujioka, a town about 90 kilometers northwest of Tokyo (then known as Edo). After centuries of internal warfare, the Tokugawa clan had unified and pacified Japan at the start of the 17th century and ruled through military commanders known as shoguns, while Japan's emperor became a figurehead. Japan's legendary samurai warriors remained a privileged elite atop a rigid, feudal class hierarchy.

No one knows who introduced Seki to mathematics. It appears he was largely self-taught, partly using Chinese texts; the shoguns' policy of isolation kept *wasan* almost free of influences from other parts of the world. Seki's mathematical proficiency landed him a job keeping accounts and mapping for a Tokugawa lord whose fief lay just west of Edo. When that lord's son became the shogun, Seki joined the shogunate in the capital.

Science historians have long believed

*International Conference on History of Mathematics in Memory of Seki Takakazu, 25–31 August 2008, Tokyo University of Science.

wasan was more of a refined accomplishment—like mastering the tea ceremony—than a serious scientific pursuit. Jochi says that was probably true during the late 18th and early 19th centuries. But for Seki and other contemporary samurai, who had to develop skills useful for a peaceful bureaucracy, knowledge of *wasan* "was a tool for success in life."

The accounting jobs apparently gave Seki the leisure to work on mathematics. Among other advances, he devised new notation for

***Wasan* master.** Seki's career proves that mathematics in feudal Japan was more than a genteel hobby.

handling equations with several variables and developed solutions for equations with an unknown raised to the fifth power. His most significant work focused on determinants, numbers that capture characteristics of matrices, a field he pioneered a year or two ahead of his European contemporary Gottfried Leibniz.

Seki published only one book during his lifetime. After his death in 1708, two of his students, the brothers Katahiro and Takaakira Takebe, gathered Seki's more erudite work into 20 volumes; another follower, Nushizumi Yamaji, published a book on Seki's methods of measuring circles and arcs and founded the Seki School of Mathematics.

Mining those posthumous texts, Komatsu

says he turned up a significant overlooked achievement: Seki's discovery around the year 1680 of a general theory of elimination, a method of solving simultaneous equations by whittling down the number of unknown quantities one by one. Komatsu says René Descartes had hinted at an elimination theory in 1627. Another French mathematician, Étienne Bézout, advanced the theory 150 years later.

Unlike Descartes, however, Seki had no real successors; his work was the high-water mark of *wasan*. After the Takebe brothers, the Japanese mathematical tradition hit a dead end. Later in the Tokugawa era, says Tatsuhiko Kobayashi, a science historian at the Maebashi Institute of Technology in Japan, texts entering the country through the tightly controlled trade with China and Holland began bringing news of advances made elsewhere in the world. The isolation policy ended in 1853. In 1868, dissident lords overthrew the Tokugawa shogunate and restored imperial rule, ushering in a modernization drive that included an 1872 governmental order to replace *wasan* with Western mathematics in school curricula.

"The demise of *wasan* stemmed largely from its divorce from the natural sciences," Mark Ravina, a historian at Emory University in Atlanta, Georgia, wrote in the Summer 1993 issue of the journal *Monumenta Nipponica*. Scientists of the Tokugawa era, Ravina wrote, considered *wasan* intriguing but useless for astronomy, calendar-making, and surveying. Rivalries among the schools that formed around *wasan* masters also hurt Japanese mathematics. Many schools had their own schemes of mathematical notation and treated problem-solving techniques as trade secrets.

More simply, Komatsu says, Seki's more erudite work "was too difficult for people to pick up and carry forward." For example, Komatsu believes Seki's elimination theory fell into obscurity because it required computations that were impossible without number-crunching computers.

Could things have turned out differently? Historians can only speculate. Seki worked on determinants simultaneously with Leibniz, another mathematician whose work went unrecognized for decades because he never published it. "There were striking similarities in mathematical thinking" between the two men, says Eberhard Knobloch, a Leibniz scholar at the Berlin University of Technology. If the Eastern and Western mathematical sages had been in contact, Knobloch says, it probably would have advanced mathematics worldwide. **–DENNIS NORMILE**

BIOETHICS

Students Learn How, Not What, To Think About Difficult Issues

A novel bioethics program trains teachers to help students confront challenges in the classroom—and in their lives

As a student, Rosetta Lee had mixed feelings about animal dissections. Now a science teacher, she gives her students the choice to opt out. That policy used to foster some unruly behavior in her classroom at Seattle Girls' School, a private middle school in downtown Seattle, Washington, where she's taught for 8 years. Those who chose not to dissect a chicken leg would often taunt their classmates with accusations of animal cruelty, whereas participants were "carelessly playful" and waved around pieces of chicken fat.

Not anymore. Thanks to a novel program that trains secondary school teachers in bioethics, Lee now prepares students with carefully guided classroom discussions exploring the potential benefits and harms of dissection. Lee can see the effect on her students: Those who choose to participate in the dissection are more studious and respectful, and those who abstain are less judgmental.

The program, sponsored by the Northwest Association for Biomedical Research (NWABR) and the University of Washington (UW), has taught Lee and other teachers how to help students think more critically about ethical issues inside and outside the classroom, such as the appropriate uses of genetic testing and the acceptability of stem cell research and genetically modified foods. That's an increasingly important skill, say science educators. "It would be a great thing for our society to have people who are more prepared to engage with these bioethical problems at that level," says Bruce Fuchs, director of the Office of Science Education at the U.S. National Institutes of Health (NIH) in Bethesda, Maryland. Teaching bioethics may even whet students' appetites

for science itself, says Jeanne Ting Chowning, the director of the NWABR program and a former high school biology teacher. "Students often ask, 'Why do we have to learn this?'" Chowning says. "This is one way to show them … why it's important to understand science."

Beyond gut feelings

Science teachers typically get very little guidance on how to approach bioethics, says Chowning. "A lot of biology textbooks give you a really interesting scenario and maybe a

Serious play. Science teacher Jamie Cooke and Mercer Island (Washington) High School students use Play-Doh to model stem cells in a developing embryo as part of a class discussion.

few discussion questions and say, 'Discuss this with your class.'" That's a terrifying prospect for many teachers, Chowning says. "They're afraid it's just going to erupt into 'my opinion versus your opinion.'"

One of the first efforts to give teachers the tools they need was funded nearly 20 years ago by Roche Pharmaceuticals. Working with

the Hastings Center, a bioethics research institute in Garrison, New York, the New Jersey Science Supervisors Association (NJSSA) developed a bioethics curriculum that includes case studies and classroom guides. Although the materials are still in use, much of it "was developed by ethicists who haven't been in a high school classroom since they themselves were in high school," says Lola Szobota, a district science supervisor in northern New Jersey who co-directs the NJSSA program and serves as an adviser to NWABR.

Chowning and colleagues wanted to build on that effort, which was restricted to New Jersey teachers. In 2003, they received a 5-year, $1.5 million Science Education Partnership Award from the National Center for Research Resources, a component of NIH, to develop a primer and other materials for teachers and run summer training workshops. Last month, they received notice of a new grant for $1.3 million, with the aim of developing additional materials and training for teachers and expanding the program's reach.

The primer (downloadable at www.nwabr.org/education/index.html) provides a step-by-step process for ethical decision-making. In Lee's class, that means helping her students identify the ethical question (to dissect or not to dissect); examine relevant facts about the planned lab, including what they might expect to learn from a dissection that they couldn't learn from a book; consider different perspectives; and weigh the possible courses of action.

Chowning and colleagues have also published curricula on stem cells and on HIV vaccine research, and there's one in the works on genetic testing for nicotine addiction risk in collaboration with UW's Department of Genome Sciences. The stem cell curriculum, for example, begins with a lesson on the biology of stem cells and a lab exercise in which students experiment with *Planaria* flatworms, whose stem cells enable them to regenerate when cut into pieces. In later lessons, students delve into how scientists obtain embryonic and adult stem cells and discuss a case study in which a couple who conceived two

children with in vitro fertilization has to decide the fate of their frozen embryos. The students finish the unit with a letter to the president with recommendations for government policy on stem cell research.

Each summer, about 25 teachers gather for a week at a rustic forest retreat on the eastern slope of Mount Rainier to practice strategies described in the ethics primer, develop case studies, and vet one another's work. "Last year, I came away from every session with something I could actually use in class, and that's not something I can say about any of the other [workshops] I've been to," says repeat attendee Tracy Watts, a teacher at Ontario High School in Ontario, Oregon. The NWABR program has also reached more than 2000 teachers through shorter workshops at national teacher conferences, Chowning says, and more than 500 teachers have downloaded materials for their Web site.

Teachers are free to adapt their training to fit their needs. Some, like Jamie Cooke, a science teacher at Mercer Island High School near Seattle, have developed entire courses on bioethics. The majority use it in smaller doses, incorporating a 2-week stem cell unit into a biology class, for example, or sprinkling bioethics lessons throughout an existing curriculum.

Cooke, who teaches in an affluent suburban district, says that bioethics appeals to college-bound students looking for a challenging science course as well as to those who just want to learn more about topics making news. At Kent-Meridian High School in Kent, Washington, an ethnically diverse and highly transient urban school where about 70% of the students qualify for government-subsidized lunches, biology teacher Jodie Mathwig uses movies and case studies as the basis for bioethics discussions that help engage students with little prior interest in science. "Stories where real people have difficult decisions to make really get the kids interested," she says. Often, Mathwig says, the students realize they need to understand biology before they can take a stand.

Like several other teachers, Lee says some parents asked her initially if she was "trying to teach values to my kid." But their concern vanished, says Lee, after she explained she was teaching students *how* to think through ethical dilemmas, not what to think about any given issue.

Meeting a need

The NWABR project is not the only effort to help teachers incorporate ethics discussions into their classes. The Kennedy Institute of Ethics at Georgetown University in

Washington, D.C., maintains a free database of case studies and other materials (at highschoolbioethics.georgetown.edu), lends out videos, and does library research to find suitable materials for any teacher who requests them.

At the University of Utah in Salt Lake City, geneticist Louisa Stark oversees a genetics education Web site (teach.genetics. utah.edu/) with lesson plans on topics such as gene therapy and personalized medicine and guides to discussing the ethical issues raised by these emerging biotechnologies. The Utah group also runs weeklong bioethics workshops for high school teachers and sessions at national conferences for science teachers.

A new project at the University of Pennsylvania (Penn) will help teachers tackle topics in neuroethics, such as potential forensic and military uses of brain-imaging technology and the care of patients in a persistent vegetative state (see hsneuroethics.org/). Funded by the Dana Foundation and led by bioethics graduate student Dominic Sisti, the program will supplement a high school bioethics project begun several years ago by Penn bioethicist Arthur Caplan. The group is developing a neuroethics primer and will run workshops for local teachers.

There seems to be both the demand and need for such efforts. Brian Shmaefsky, a board member of the National Association of Biology Teachers, says he's noticed a steady rise in the number of proposed bioethics workshops at NABT's annual meeting, as well as growing attendance at those workshops. Science standards released in 1996 by the U.S. National Academies make no explicit mention of bioethics, but they assert, for example, that students "need to take informed positions on some of the practical and ethical implications of humankind's capacity to manipulate living organisms." Most state standards include similar language, Chowning and others say.

Quantifying the impact of bioethics in the classroom is difficult, says Carolyn Cohen, a

Fun with ethics. Jodie Mathwig (*left*) and Jacob Dahlke review lesson plans at a summer bioethics workshop for teachers.

Seattle-based program evaluator. In a program assessment completed last month, Cohen found that about 90% of the teachers surveyed believe their students have a better understanding of the role of science in society as a result of the bioethics lessons they've received and that nearly 80% reported that students have become more aware of differing points of view. Slightly more than 60% reported improvements in students' scientific literacy and understanding of how scientific research is conducted. The teachers themselves feel better able to incorporate bioethics into their classes and do so more frequently. The new NIH grant includes funding for a more rigorous evaluation of students who have been taught bioethics and a comparable group that has not.

Chowning says the NWABR program needs to become self-sustaining to make a lasting impact on educational practices. Toward that end, she has been developing an online version of the workshop—complete with video of experienced teachers in the classroom, voice-over PowerPoint presentations, and other multimedia components—that teachers can enroll in through UW as part of ongoing professional development.

In the meantime, she hopes that students are acquiring the skills to make better decisions as adults. "They'll have to confront issues around genetic testing in their families, and they'll have to vote on issues like embryonic stem cell research," she says. "We need to make sure they're prepared for their future."

–GREG MILLER

COMMENTARY

Illustrating natural science	Fish plates and fitness	Species responses to climatic shifts
196	204	206

LETTERS | BOOKS | POLICY FORUM | EDUCATION FORUM | PERSPECTIVES

LETTERS

edited by Jennifer Sills

Declines in NIH R01 Research Grant Funding

IN ASSOCIATION WITH THE NATIONAL CAUCUS of Basic Biomedical Science Chairs, we have tracked funding of R01 grants (1–3). We found an R01 decline, which slows progress in fundamental research and deters bright young people from entering science.

In 2006, we reported (3) that between fiscal years (FY) 1999 and 2005, there was a sharp decline in the funding of unamended (i.e., as originally submitted) R01 applications. In the past 2 years, further declines have occurred for both new (Type-1) and renewal (Type-2) R01 applications (Table 1) (4, 5).

New data for FY2007 indicate a substantial drop in the number of R01 applications submitted. However, during FY2006 and FY2007, further decreases occurred in total number of unamended applications funded and in dollar allocations. The 7% success rate of new applications implies that only 1 of 14 unamended applications was funded. For specific NIH Institutes, such as the National Cancer Institute, National Institute of Allergy and Infectious Diseases, and National Institute of Neurological Disorders and Stroke, success rates were even lower: 5%, 5%, and 3%, respectively. For renewal applications, the decline means discontinuation of 75% of ongoing programs.

Resubmission of amended applications—a slow, time-consuming process—increases likelihood of success (6) but protracts initiation of research. For ongoing projects (Type-2 applications), interruption of in-progress investigations often breaks up successful, experienced teams of investigators.

We also calculated annual allocations to the entire R01 program in relation to fluctuations in total NIH Research grant support (7). Since FY2000, R01 funding has suffered compared with overall funding, so that by FY2007 the deficiency reached almost $1.2 billion (Table 2). Rectification of this progressive decline in

R01 funding would provide about 3200 additional research grants. Selective de-emphasis of R01 grants limits innovative discoveries for improving our nation's health.

H. GEORGE MANDEL[1]* AND ELLIOT S. VESELL[2]

[1]Department of Pharmacology and Physiology, The George Washington University School of Medicine and Health Sciences, Washington, DC 20037, USA. [2]Department of Pharmacology, Pennsylvania State University College of Medicine, Hershey, PA 17033, USA.

*To whom correspondence should be addressed. E-mail: phmhgm@gwumc.edu

References and Notes

1. H. G. Mandel, E. S. Vesell, *Science* **294**, 54 (2001).
2. H. G. Mandel, E. S. Vesell, *J. Clin. Invest.* **114**, 872 (2004).
3. H. G. Mandel, E. S. Vesell, *Science* **313**, 1387 (2006).
4. Kindly supplied by the Division of Information Services, Office of Research Information Services, NIH.
5. The new data differ from those previously published because under unsolicited grants, NIH now combines R01 data with Program Announcements (P.A.'s). Success rates for P.A.'s did not differ appreciably from those for R01's, at least through FY1999 to 2005 when data for P.A. success rates were reported separately. For FY2004 and 2005, P.A.'s represented 20.4% of the combined pool, and for the 5 prior years the proportion was 15.6%.
6. For FY2007, first-time and second-time revisions have provided funding for an additional 1573 and 1272 grants, and $321.1 and $470.5 millions for new grants. For Type-2 amended applications, these numbers are 932 and 626, and $352.5 and $228.3 millions, respectively.
7. These figures include Type-1 and Type-2 grants, competing, noncompeting, and supplements; http://grants1.nih.gov/grants/award/research/Research_by_Activity_Code.xls.

A Call to Action for Coral Reefs

AT THE 11TH INTERNATIONAL CORAL REEF Symposium (ICRS) held in July in Fort Lauderdale, Florida, midway through the International Year of the Reef, more than 3500 experts from 75 countries assembled to face some hard truths: Coral reefs are teetering on the edge of survival, and it is our fault. High levels of carbon dioxide in the atmosphere can produce a lethal combination of warmer seawater and lower pH. Pervasive

Table 1. Fate of unamended, unsolicited "R01 Equivalent"* research grant applications (4)

FY	Number reviewed	Number awarded	Total $, millions awarded	Success rate %
Type-1 grants (new submissions)				
2000	10,284	2084	616.1	20.3
2001	9851	1864	599.6	18.9
2002	10,083	1831	617.5	18.2
2003	11,511	1733	587.6	15.1
2004	13,370	1595	551.8	11.9
2005	13,578	1236	443.9	9.1
2006	13,659	941	332.5	6.9
2007	12,021	864	321.1	7.2
Type-2 grants (renewal submissions)				
2000	3374	1787	589.9	53.0
2001	3218	1687	598.8	52.4
2002	3270	1614	582.7	49.4
2003	3922	1765	654.9	45.0
2004	3955	1606	613.2	40.2
2005	4128	1335	525.3	32.3
2006	3881	998	389.0	25.7
2007	3605	909	372.9	25.2

* "R01 equivalents" include a small number of R29 and R37, as well as P.A. grants.

Table 2. Progressive, steady decline in R01 grant funding during the past 8 years in relation to total NIH research grant support (7)

FY	Total NIH research grants	Actual R01* funds allocated	Calculated R01 fund, in proportion	Difference between calculated and actual funding (in millions)
2000	13,003	7141	7141	0
2001	14,908	8093	8187	−94
2002	16,830	8985	9243	−258
2003	18,461	9742	10,138	−396
2004	19,608	10,176	10,768	−592
2005	20,206	10,288	11,097	−809
2006	20,154	10,122	11,068	−946
2007	20,416	10,046	11,212	−1166

*These numbers are actual R01 grants and are exclusive of R29 and R37 grants.

overfishing, pollution, coastal development, and physical damage further undermine reef health—and consequently the health of the people and ecosystems depending upon them (*1*).

Coral reefs feed, protect, and provide livelihoods for hundreds of millions of people around the world. They create homes for billions of fish and other animals, buffer coastlines from the ravages of storms, and provide rich economic opportunities through tourism and fishing. Their value to society has been estimated at more than $300 billion per year. Reefs are the dynamic centers of the most concentrated biodiversity on Earth.

It is not too late to save coral reefs. A consensus emerged at the 11th ICRS that society has both the knowledge and the tools to bring coral reefs back from the brink. We have a real—but rapidly narrowing—window of opportunity in which to take decisive action. We must immediately: (i) Cut CO_2 emissions by lowering our carbon footprint and ask our policy-makers to commit to low carbon economic growth. (ii) Eliminate open-access fisheries in coral reef ecosystems and instead establish and enforce regulations on user rights, total allowable catch, individual catch quotas, nondestructive gear, and other sustainable fisheries regulations. (iii) Protect coral reef herbivores, including parrotfish, by banning the harvesting of these species for sale and commercial consumption. (iv) Establish and strictly enforce networks of Marine Protected Areas that include No-Take Areas. (v) Effectively manage the waters in between Marine Protected Areas. (vi) Maintain connectivity between coral reefs and associated habitats; mangroves, sea grass beds, and lagoons contribute to the integrity of reef ecosystems and their continued production of ecosystem services. (vii) Report regularly and publicly on the health of local coral reefs. (viii) Recognize the links between what we do on land and how it affects the ocean. (ix) Bring local actors together—including members of industry, civil society, local government, and the scientific community—to develop a shared vision of healthy reefs and a road map for getting there.

Only by taking bold and urgent steps now can we hope to ensure that reefs will survive to enrich life on Earth, as they have for mil-

LIFE IN SCIENCE

Sounds of Atoms

Early in setting up our nanoscience laboratory at Penn State, we were frustrated because we could not peer into the tunneling junctions of our scanning tunneling microscopes (STMs) to see what the atoms were doing. We were particularly vexed when singular events, such as an atom moving

on the end of the STM probe tip, confounded our data and forced us to start over. Such an event was very difficult to identify; it would be just a blip in a recorded image or a flash on an oscilloscope, and it would not be recognizable at all in the frequency spectrum of noise monitored on our spectrum analyzer.

The solution to our problem was right there in the mirror every morning but had nothing to do with sight. Humans can hear a wide dynamic range and have a fantastic ability to recognize patterns in sounds. Borrowing a trick from electrophysiology, we sent the tunneling current of our microscopes into an audio amplifier and then turned our probe tip height into a tune by applying that signal to a voltage-controlled oscillator. One stereo channel corresponds to tunneling current and the other to STM tip height. This way, we could hear when imaging was proceeding well and when something was amiss. We applied this approach to every aspect of our microscopy and spectroscopy. The dreaded sound of an atom moving on the tip of a low-temperature STM has a characteristic rising "whoook" that is unmistakable (and sometimes heartbreaking). Every student, postdoc, and visitor in the group became familiar with the songs of his or her microscope—they learned which tune meant which problem and always rejoiced when they heard the best melody of all: the sound of a happily working microscope.

Such ideas were also developed independently at the IBM Zurich and IBM Almaden Laboratories, and elsewhere. If you would like to hear some of the sounds of our microscopes and of imaging molecules, please listen to the media files in the supporting online material (*1*).

PAUL S. WEISS[1]* AND STEPHAN J. STRANICK[1,2]

[1]The Pennsylvania State University, University Park, PA 16802, USA. [2]National Institute of Standards and Technology, Gaithersburg, MD 20899, USA.

*To whom correspondence should be addressed. E-mail: stm@psu.edu

References and Notes
1. Some of the sounds of our microscopes, recorded by P. Han, A. Kurland, and P. S. Weiss of Penn State, are available as supporting material on *Science* Online (www.sciencemag.org/cgi/content/full/322/5899/190a/DC1).
2. Our original work was supported by the National Science Foundation, the Office of Naval Research, the Petroleum Research Fund administered by the American Chemical Society, and G. Marcotrigiano of Gary's Electronics.

EDITOR'S NOTE

This will be an occasional feature highlighting some of the day-to-day humorous realities that face our readers. Can you top this? Submit your best stories at www.submit2science.org.

lions of years before us. By failing to act, we risk bequeathing an impoverished ocean to our children and future generations (*2*).

RICHARD E. DODGE,[1]* CHARLES BIRKELAND,[2] MAREA HATZIOLOS,[3] JOAN KLEYPAS,[4] STEPHEN R. PALUMBI,[5] OVE HOEGH-GULDBERG,[6] ROB VAN WOESIK,[7] JOHN C. OGDEN,[8] RICHARD B. ARONSON,[9] BILLY D. CAUSEY,[10] FRANCIS STAUB[11]

[1]Oceanographic Center, Nova Southeastern University, Dania Beach, FL 33004, USA. [2]Department of Zoology, University of Hawaii, Honolulu, HI 96822, USA. [3]The World Bank, Washington, DC 20433, USA. [4]Institute for Study of Society and the Environment, National Center for Atmospheric Research, Boulder, CO 80307, USA. [5]Hopkins Marine Station, Pacific Grove, CA 93950, USA. [6]Centre for Marine Studies, University of Queensland, Brisbane, QLD 4072, Australia. [7]Department of Biological Sciences, Florida Institute of Technology, Melbourne, FL 32901, USA. [8]Florida Institute of Oceanography, St. Petersburg, FL 33701, USA. [9]Dauphin Island Sea Lab, Dauphin Island, AL 36528, USA. [10]National Marine Sanctuary Program, Key West, FL 33040, USA. [11]International Year of the Reef Coordinator, Bethesda, MD 20814, USA.

*To whom correspondence should be addressed. E-mail: dodge@nova.edu

References and Notes
1. An overview of the 2632 papers presented can be found on www.nova.edu/ncri/11icrs/outcomes.html.
2. Please add your name to sign up for this Call to Action. Go to www.thepetitionsite.com/1/11th-international-coral-reef-symposium-call-to-action.
3. R.E.D. is the Chair of the 11th ICRS Local Organizing Committee (LOC); C.B., M.H., J.K., S.R.P., and O.H.-G. are Super Chairs of the 11th ICRS Mini-Symposia. R.v.W. is the Science Chair of the LOC. J.C.O. and B.D.C. are LOC organizers, R.B.A. is the President of the International Society for Reef Studies, and F.S. is Coordinator of the International Year of the Reef 2008.

CREDIT: PETER HOEY
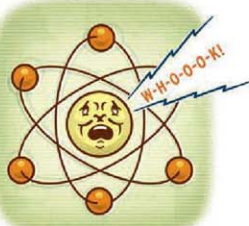
Neutralizing the Impact Factor Culture

IN THE LETTER "PAINFUL PUBLISHING" BY M. Raff *et al.* (4 July, p. 36) and in the accompanying Editorial "Reviewing peer review" by B. Alberts *et al.* (4 July, p. 15), the authors succinctly outline the pressure being felt by both junior and senior scientists to publish in high-profile journals. Although not defined by the authors, high-profile journals are generally identified by and have become synonymous with Thomson high–impact factor scores. A common, but deeply flawed, practice has been to equate the importance and quality of a paper with the impact factor score of the journal in which it is published. In many cases, decisions on obtaining jobs, seeking tenure and receiving promotions and grants are being based on the impact factor of the journals in which an individual publishes. This creates enormous pressure to publish in high–impact factor journals. This situation has become so extreme that in some institutions the impact factor of each published paper in a scientist's bibliography is being requested and/or checked, junior scientists have become reluctant to initiate experiments that may not lead to publication in high–impact factor journals, and candidates for certain positions are being told that their chances are slim if they don't have papers in *Science*, *Nature*, or the like. As a result, many scientists are now more concerned about building high–impact factor bibliographies than their science.

The adverse effects of the impact factor culture must be reversed before more damage is done to the orderly process of scientific discovery. Although there may be no way of stopping computer-generated evaluation of journals and published papers, the scientific community certainly can control its use. To accomplish this, several concrete steps should be taken. First, each institution should make it clear, in a written statement, that it will not use the impact factor or the like to evaluate the contributions and accomplishments of its staff. Second, the heads of laboratories should prepare similar written statements and in addition discuss in depth with their fellows the importance of solid step-by-step science. Third, the editors of journals published by professional societies, joined by as many other journal editors as are willing, should indicate that they will not advertise, massage, or even state the impact factor score of their respective journals. By means such as these, it might be possible to put science back on the right track.

ABNER L. NOTKINS

Experimental Medicine Section, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD, USA. E-mail: anotkins@mail.nih.gov

Note
1. The views expressed here do not represent the position of the NIH.

Impact Factor Fever

IN A RECENT EDITORIAL ("REVIEWING PEER review," 4 July, p. 15) B. Alberts *et al.* addressed the most important problem affecting the scientific community today: the incredible pressure to publish, which is the drift of the "publish or perish" philosophy. Scientific quality is bound to suffer when scientists focus only on their publication records.

As an author, reviewer, and editor of a small international scholarly journal, I have noticed a dramatic increase in plagiarism, "salami-slicing" science, and other kinds of research misconduct over the past few years.

I fully agree that the peer-review process should be revised in order to reduce its length and make it less agonizing for authors, reviewers, editors, and readers (*1*). Some of the methods suggested in the Editorial, such as sending reviews on to other journals and enlarging the pool of referees, are certainly needed and will hopefully be successful. However, Alberts *et al.* failed to mention what is perhaps the most debilitating illness plaguing the scientific community, which I call the "impact factor fever." The exacerbated pressure to publish we all suffer from is induced by an exaggerated reverence for the impact factor.

Scientific achievement cannot be soundly evaluated by numbers alone. As Albert Einstein reputedly said, "Not everything that can be counted counts, and not everything that counts can be counted." How long must we wait until an antidote against the impact factor fever is developed?

PAOLO CHERUBINI

Dendrosciences, WSL Swiss Federal Research Institute, CH-8903 Birmensdorf, Switzerland.

Reference
1. M. Raff, A. Johnson, P. Walter, *Science* **321**, 36 (2008).

LETTERS

CORRECTIONS AND CLARIFICATIONS

ScienceScope: "Aussie science review" by E. Finkel (19 September, p. 1619). John Mattick is no longer director of the Institute for Molecular Bioscience. His correct title is Professor of Molecular Biology and ARC Federation Fellow at the Institute for Molecular Bioscience, University of Queensland.

Editors' Choice: "Adding less or substrating more?" (29 August, p. 1134). The penultimate word in the final sentence should have been "extinction," not "speciation." The final sentence should read, "Their results also show that an apparent excess of recently diverged lineages in lineage-through-time plots ... can be produced when declining net diversification is driven by increasing extinction rates." Also, the first author of the *Evolution* study was Rabosky, not Rabolsky.

Reports: "Limits for combustion in low O_2 redefine paleoatmospheric predictions for the Mesozoic" by C. M. Belcher and J. C. McElwain (29 August, p. 1197). In Fig. 2, the data point positioned at 250 Ma in the series labeled Falkowski *et al.*, 2005 is incorrect; it should have been placed at 205 Ma (see corrected figure below). There were several references to this in the text that need to be amended: The fourth sentence of the first paragraph on page 1197 should read "Few proxies have been developed for testing past atmospheric O_2 concentrations, particularly the low levels inferred for the Permo-Triassic (4) and the Jurassic (2, 3)." The fourth sentence of the second paragraph on page 1199 should read "This analysis revealed that wildfires were prevalent throughout the Mesozoic and, coupled with data from our combustion experiments, did not support model-based predictions of low O_2 (<15%) for the Jurassic (2, 3) (Fig 2)." The sixth sentence of the third paragraph on page 1199 should read "On the basis of our newly proposed low limit for combustion, both the Falkowski *et al.* (3) and GEOCARB-SULF (2) models are currently incompatible with the record of fires in the Mesozoic, because they predict extensive periods of low (10 to 12%) atmospheric O_2 throughout the Jurassic (2, 3)." This error does not affect the results or conclusions of this Report.
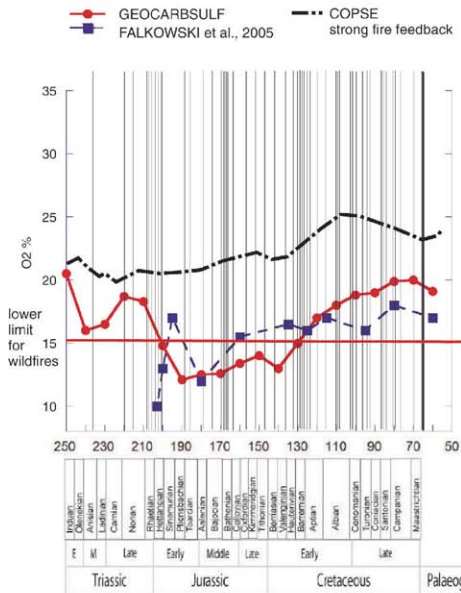
Technical Comments: Response to Comments on "A semi-empirical approach to projecting future sea-level rise" by S. Rahmstorf (28 September 2007, p. 1866; www.sciencemag.org/cgi/content/full/317/5846/1866d). It was stated that the semi-empirical formula for projecting sea-level rise can successfully predict the second half of the sea-level data when trained only on the first half of the data. This is correct, but it was illustrated by an incorrect figure (Fig. 1), in which the first half of the smoothed sea-level curve (1882 to 1941) was used to predict the sea level for 1942 to 2001. Because the smoothing procedure used a 15-year time window, the smoothed sea-level curve up to 1941 effectively contains sea-level information up to 1948. When this error is corrected and only annual sea-level measurements from 1882 to 1941 are used, the obtained fit gives a sea-level slope of 0.35 mm/year per °C, and the base temperature is −0.46°C. This is in fact closer to that obtained using the full data, and the sea-level prediction for 1942 to 2001 is within 1.4 cm of the (15-year smoothed) observed sea level (the Response stated that it is within 2 cm). No conclusions in the Response are changed by this correction.

COGNITIVE SCIENCE

Arguing for Embodied Consciousness

Harold Fromm

To describe his project, Edward Slingerland could hardly have chosen a more direct title than *What Science Offers the Humanities: Integrating Body and Culture*. Coming from an academic career in Asian studies and theology (as a nonbeliever), Slingerland (a co-founder of the Centre for the Study of Human Evolution, Cognition, and Culture at the University of British Columbia) prepared by spending the past five years reading widely and deeply in cognitive neurosciences and the philosophy of science. Although he mainly focuses on consciousness, his overall task is to address the befuddled dualism that still dominates most of our intellectual disciplines.

Slingerland's central theme is that everything human has evolved in the interests of the materiality of the body. He identifies objectivist realism and postmodern relativity, both insufficiently attentive to the body, as the major epistemologies to be swept away, followed by the dualism of body and soul. For Slingerland, the presiding genii behind such a cleansing are George Lakoff and Mark Johnson, with heavier debts to Johnson [whose terse summary of embodiment in (*1*) appeared too late for Slingerland to reference]. They view all thought and human behavior as generated by the body and expressed as conceptual metaphors that translate physical categories (such as forward, backward, up, and down) into abstract categories (such as progress, benightedness, divinity, immorality). These body-driven metaphors, Slingerland writes, are a "set of limitations on human cognition, constraining human conceptions of entities, categories, causation, physics, psychology, biology, and other humanly relevant domains."

The supposedly objective world is not "some preexisting object out there in the world, with a set of invariant and observer-independent properties, simply waiting to be found the way one finds a lost sock under the bed." All we can ever see or understand is what our own bodily faculties permit via the current structure of the brain.

In opposition to objective realism, postmodern relativity regards language and culture as constituting the only "real" world possible for us. It posits an endless hall of mirrors with no access to outside—epitomized by Derrida's notorious remark that there is nothing (at least for humans) outside of texts (i.e., culture). This view, which dominated the humanities for several decades, is mercifully beginning to fade as the cognitive sciences have matured and are increasingly promulgated.

Even though the knowing human subject is itself just a thing and not an immaterial locus of reason, the universe it experiences is as real

WHAT IF WE JUST THINK WE'RE ROBOTS?
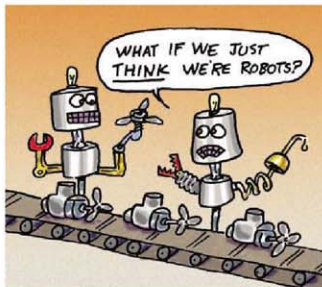
and functional for us as any "thing" could possibly be. We do get some things "right," even if we can never know the noumenal genesis behind our knowledge. And the very concept of noumena (things in themselves independent of any observer) now seems somewhat obsolete, given that the intuition of discrete, self-bounded "things" is as built-in to the human psyche as the Kantian intuitions of space and time, grounding all experience.

Our million billion synapses produce a "person" with the illusion of a self. Slingerland holds that "we are robots designed to be constitutionally incapable of *experiencing* ourselves and other conspecifics as robots." Our innate and overactive theory of mind (that other people, like ourselves, have

"intentions") projects agency onto everything—in the past, even onto stones and trees. The "hard problem" for philosophy of consciousness (to use David Chalmers's phrase) remains: what are thoughts, cogitations, thinkers, qualia? Chalmers's solution, alas, swept away Cartesian dualism only to sneak his own magic spook, conscious experience (for him, on par with mass, charge, and space-time), in through the back door (*2*, *3*).

Slingerland starts with Darwin and eventually follows Daniel Dennett so far as to agree that consciousness can be done full justice through third-person descriptions that require no mysterious, unaccounted-for, nonmaterial, first-person entity as substrate. Thus the famous "Mary," who intellectually knows everything there is to know about color despite having been sequestered for life in a color-free lab, will recognize red the first time she steps outside (*4*). And Thomas Nagel's famous bats don't know anything about bathood that we can't figure out for ourselves from observation (*5*). No first-person construct, no locus of consciousness, need be invoked.

The next step, if you want to go so far (the jury is out), is to eliminate consciousness altogether, because there's nothing for it to do that can't be done without it. And with it, you need a spook to keep the show on the road. Choose your insoluble problem: eliminate consciousness altogether as superfluous or explain it (if there's really a you who makes such choices). Slingerland prefers the first option.

His conclusion, which I can hardly do justice to here, is relatively satisfying. He notes that although we don't have great difficulty knowing that Earth revolves around the Sun while feeling that the Sun is rising and setting (Dennett's favorite example of folk psychology), "no cognitively undamaged human being can help *acting* like and at some level really *feeling* that he or she is free"—however nonsensical the notion of agencyless free will (i.e., "choices" without a self to make them). Still, once the corrosive acid of Darwinism [to use Dennett's figure from (*6*)] has resolved the body-mind dualism into body alone, some but not most of us are able "to view human beings simultaneously under two descriptions: as physical systems and as persons."

What does Slingerland's intellectually acute, wide-ranging, well-written, and deeply knowledgeable survey of the hard and soft disciplines behind consciousness achieve? For some from the humanities, this book will be a

The reviewer is at the Department of English, University of Arizona, Tucson, AZ 85721, USA. E-mail: hfromm@earthlink.net

CREDIT: JOE SUTLIFF

revelation. But for those clinging to cultural construction or belief in an immaterial self, still needed is a scaled-down version of *What Science Offers the Humanities* that reduces its length by half and removes the various byways and densities that might dim the epiphanal light bulbs needed to initiate conversion experiences. And the book (in whichever form) wouldn't do the hard sciences any harm either.

References

1. M. Johnson, *The Meaning of the Body: Aesthetics of Human Understanding* (Univ. of Chicago Press, Chicago, 2007).
2. D. J. Chalmers, *J. Consciousness Stud.* **2**, 200 (1995).
3. D. J. Chalmers, *The Conscious Mind: In Search of a Fundamental Theory* (Oxford Univ. Press, Oxford, 1996).
4. F. Jackson, *Philos. Q.* **32**, 127 (1982).
5. T. Nagel, *Philos. Rev.* **83**, 435 (1974).
6. D. C. Dennett, *Darwin's Dangerous Idea: Evolution and the Meaning of Life* (Simon and Schuster, New York, 1995).

SCIENCE AND ART

Subject Matter Matters

Mary Parrish

Having received a bachelor of fine arts degree at a time when artists were trained to look beyond realism for inspiration, that subject matter was not a criterion on which art should be judged, and that illustration is not art (and then later becoming a scientific illustrator), I read with interest Darryl Wheye and Donald Kennedy's *Humans, Nature, and Birds* and Jane Davidson's *A History of Paleontology Illustration*. The books offer interesting and well-constructed overviews of the natural history and imagery of birds and fossils. At the same time, the authors use their material to support the shared thesis, outlined in their introductions, that the science found in the images unites the material into a unique genre of art.

Humans, Nature, and Birds is an outgrowth of Stanford University's online Ornithological Artist Registry of bird art (*1*). Selections from the book, including sample images, have been posted online for public access (*2*). Wheye (an artist and writer) and Kennedy (a former editor-in-chief of *Science*) present birds as a case study to describe what they call "Science Art," assigning the term to a

The reviewer is at the Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20013–7012, USA. E-mail: parrishm@si.edu

genre of art that contains scientific content. In their view, Science Art needs to say "something about the natural world and how it works" and should always be "accompanied by an explanatory caption that helps the viewer decode the underlying science."

Their text highlights the science they find in (and behind) the images they selected, and I found their explanations enriched my understanding of the material. Further, I appreciated Wheye and Kennedy's awareness that there is often more to the creation of the images than what immediately meets the eye. Commonly, at least 85% of the effort in preparing a scientific illustration is spent on (often collaborative) research, and only some 15% on visual rendering of the findings from that research. A seemingly simple line drawing such as the Cretaceous landscape (Davidson's figure 6.7) that Edward Vulliamy prepared for Albert

Minimal realism. Charley Harper reduced his subjects, such as *Eskimo Curlew* (serigraph, 1957), and their ecological context to elementary visual terms.

Charles Seward (*3*) may have required reconstructing each species from fragmentary material and descriptions (possibly contradictory) by several paleobotanists. I also enjoyed gaining a glimpse into what some scientists see when they look at art.

Even when the focus is restricted to avian examples, Science Art covers a wide spectrum. The book features some 60 beautifully reproduced color images that collectively span 30,000 years of birds in art. These include an owl from Chauvet Cave in France (the oldest known cave painting that depicts a bird); Pieter

---

Brueghel the Elder's *The Peasant and the Nest Robber* (1568), with its "explicit conservation message"; Paul Klee's expressionist painting *Twittering Machine* (1922), which some have interpreted as a satire of laboratory science; and *KEO the Winged Satellite*, the passive satellite time capsule envisioned by Jean-Marc Philippe (1998).

Humans, Nature, and Birds resembles a catalog for an art exhibition. It is divided into three major sections: a Lower Gallery ("Bird Art over the Millennia"), a Mezzanine, and an Upper Gallery ("How Science and Art Overlap"). The two galleries each contain five thematic rooms filled with images. The art appears on the left pages, opposite explanatory text that features scientific interpretations.

The Mezzanine contains no images but offers a very interesting essay on aesthetics as it relates to science and art along with discussions about the interplay of the cultures of science and art and the work processes of scientists and artists. The authors suggest (à la Denis Diderot) that an aesthetically profound image records an emotional response to a rational idea. Text illuminates art; art illuminates text (*4*).

A History of Paleontology Illustration stems from a course of the same name that Davidson, a professor of art history, teaches at the University of Nevada, Reno. She is a specialist in 17th-century Dutch and Flemish painting who also teaches courses in the history of science and has a passion for paleontology, and these themes meld well in the book.

The 82 black-and-white figures and 8 color plates (the majority of which depict vertebrate material) representing 500 years of fossil imagery emerge from the matrix of paleontology itself. Davidson structures the book around a chronological history of paleontology, starting in the 15th century and continuing to the present decade.

Davidson's material ranges from the beautiful, such as J. C. McConnell's detailed drawings of crinoid specimens (1883), to the

unusual, such as the fossil amphibian Louis Bourguet discussed (5) as *Homo diluvii testis* ("the man who witnessed the Flood"). The illustrations were prepared in a variety of media, including oil paint, pencil, ink, bronze sculpture, and photography. They include depictions of individual fossil specimens, plates of fossils from several phyla, and life reconstructions of extinct animals shown individually or in their environments. The author also mentions the printing techniques (e.g., woodcuts, wood and copper engravings, and stone lithography) used to reproduce the images in their original publications.

Davidson limits her examples primarily to materials used to illustrate paleontological texts. She highlights the appearance of fossils in art as they were understood and used at the time in which the image was created. The uses of fossils in images change along with scientists' understanding of them: from objects with no organic affinities; to objects of curiosity; to the inorganic remains of once-living organisms; to objects that illuminate scientific ideas about evolution, extinction, and ancient ecology.

Unfortunately, *A History of Paleontology Illustration* is handicapped by the quality of paper and the reproduction of images, both of which are much better in *Humans, Nature, and Birds*. In addition, Wheye and Kennedy's

Capturing color. Auguste Sonrel's *Cladodus* (fossil shark) teeth are among the chromolithographs Louis Agassiz used to illustrate *Recherches sur les poissons fossiles* (6).

advice about captions would have served Davidson well. Often, both her captions and the text lack information that would be useful in identifying and understanding the illustrated material. (For example, she includes a figure from work by O. C. Marsh without naming the bones depicted or noting that they are *Triceratops*.)

Natural history and natural history images are clearly things that Wheye, Ken-

nedy, and Davidson care deeply about. But why elevate the material seen in *A History of Paleontology Illustration* and *Humans, Nature, and Birds* to a genre of art? A paper conservator once taught me that presentation (i.e., the way in which materials are housed and cared for) is the first step toward preservation. Perhaps as art, natural history images are more likely to be preserved. Further, as art, perhaps they can be used as a tool to help preserve the natural world itself. In any event, the integration of the aesthetics of art and the subject matter of science results in the elevation of both fields.

References and Notes
1. The registry was produced by Wheye in association with Kennedy, Paul R. Ehrlich, and Makoto J. Tsuchitani; see http://artist-registry.stanford.edu/.
2. www.scienceart-nature.org.
3. A. C. Stewart, *Plant Life Through the Ages: A Geological and Botanical Retrospect* (Cambridge Univ. Press, Cambridge, 1931).
4. Interested readers may enjoy Stephen Jay Gould's essay "Church, Humboldt, and Darwin: The tension and harmony of art and science" (7).
5. L. Bourguet, *Traité des petrifications* (A. C. Braisson, Paris, 1742).
6. L. Agassiz, *Recherches sur les poissons fossiles* (H. Nicolet, Neuchâtel, Switzerland, 1833–1845).
7. S. J. Gould, in F. Kelly, *Frederic Edwin Church* (National Gallery of Art and Smithsonian Institution Press, Washington, DC, 1989), pp. 94–107.

10.1126/science.1165966

BROWSINGS

Guild of Natural Science Illustrators Annual Members' Exhibit.
Ithaca, NY, July 2008.

The Guild (www.gnsi.org) met for a week of presentations, discussions, and workshops—some focused on techniques and media, others on subject matter. In conjunction with the conference, it mounted a public exhibit at Cornell University. Attendees awarded three prizes: best digital,

to Frank Ippolito's life-size *That Little Dinosaur: Mahakala omnogovae*; best traditional black & white, to Frances Fawcett's *Three Bees*; and best traditional color, to Michael Rothman's *Cycad Pollination Habitat Group* (below). This acrylic painting of Mexican taxa depicts *Zamia furfuracea*, which is pollinated by a small snout weevil, *Rhopalotria mollis*. The scene also includes the endangered cycads *Ceratozamia latifolia* (with russet leaf flushes) and *Dioon spinulosum* (with an arborescent habit).

GENETICS

Trends in Human Gene Patent Litigation

Christopher M. Holman

Fears surrounding human gene patents have, for the most part, yet to manifest themselves in patent litigation.

Critics have long warned that patents claiming human genes pose a substantial threat to public health and the progress of science (1–3). Much of the focus has been on the alleged detrimental impact of gene patents on the development and availability of diagnostic testing (1, 3, 4). Some have postulated that a "thicket" of patents will impede basic biomedical research and will stifle development and utilization of technologies that involve the use of multiple genetic sequences; DNA microarrays are a prime example (5, 6). Others claim that gene patents are uniquely difficult to design around and, thus, fundamentally more restrictive of follow-on developments than "traditional" patents (6).

In response to the perceived threat of gene patents, Congress is considering legislation

Litigation Outcomes

Outcomes for the 31 identified human gene patent litigations (24). One of the appealed cases is still pending, the others were pursued to final judicial decisions.

that would prospectively ban the patenting of not only human genes, but any "nucleotide sequence, or its functions or correlations, or the naturally occurring products it specifies" (7). Nevertheless, some argue that the problem of human gene patents and the alleged patent thicket have been overstated (1, 8), because there is little empirical evidence that these patents have had a substantial negative

University of Missouri–Kansas City, Kansas City, MO 64110, USA. E-mail: holmancm@umkc.edu

impact on research or the availability of diagnostic testing (1).

In 2004, Jensen and Murray identified 4270 U.S. patents claiming at least one human gene and concluded that one-fifth of known human genes were claimed in a patent (9). This figure raised concerns in the minds of many and was cited as justification for the proposed bill to ban the patenting of DNA (10). Other researchers have conducted surveys to gauge the chilling effect of biomedical and human gene patents on research (11) or diagnostic testing (3).

However, research on judicial enforcement is lacking. Few human gene patents have ever been asserted in court, so any chilling effect arises primarily from a perception of risk that may not comport with reality. A patent generally has no legal effect until successfully asserted in court, and attempts at judicial enforcement often fail.

To address this gap, I conducted a study to identify all instances in which a human gene patent was asserted in an infringement lawsuit and to track the results of these litigations (12–14). The foundation of this study was a search of Lexis and Westlaw databases for any U.S. patent satisfying two criteria: (i) the patent includes the canonical term "SEQ ID NO." in its claims, or includes within its claims or abstract any one of the terms used to identify a patent for inclusion in the Georgetown DNA Patent Database; and (ii) a notice of litigation has been filed indicating that the patent has been the subject of a lawsuit. Alternate search strategies were also used, including a search for any reported judicial decision involving an allegation of patent infringement and containing one of the Georgetown database terms (13). Asserted patents, complaints, and other documents generated by each lawsuit were analyzed to assess whether it involved an allegation of infringement of a human gene patent (14).

Frequency of Litigation

I identified 31 human gene patent litigations, dating back to 1987 (13, 14) (see chart, left).

Considering the large number of human gene patents (9), the substantial amount of patent litigation that has taken place involving biotechnology patents other than human gene patents (15), and the high level of concern that has been expressed with respect to the negative impact of human gene patents, 31 seems a relatively small number. For example, since 2000 at least 1294 lawsuits have been filed asserting drug patents, and 278 involving molecular biology or microbiology patents (16). Furthermore, rather than increasing, the number of human gene patent litigations pending at any given point in time has fallen off in recent years (see chart, page 199). This decline corresponds to reports of a similar marked decline in the filing (17) and issuance (18) of DNA patents in the United States since 2001.

Only 7 of the 31 lawsuits I identified involved patents identified by Murray and Jensen, indicating that their automated search strategy actually missed many human gene patents (13). However, what is most striking is that not one of the 4270 patents in their data set has ever resulted in a decision favoring the patent holder (13). Five of the cases settled before any substantive judicial decision (i.e., prior to any decision addressing the merits of the case), one was dismissed by the court, and in one litigation, the accused infringer prevailed.

Litigation Outcomes

Litigation outcome tends to vary depending on the nature of the alleged infringement (see chart, left). The vast majority of human gene patent litigations (as is the case with patent litigations in general) are dismissed before a final court decision, often as the result of the parties' reaching a settlement agreement. Six of the 16 therapeutic protein cases were litigated to a final, unappealable judicial decision (a seventh is currently on appeal), with the patentee prevailing in two cases and losing in the other four. In contrast, the litigations that did not involve a therapeutic protein all either settled or were dismissed before the case could be appealed to a higher court. Overall, lawsuits involving therapeutic proteins rarely settled before at least one substantive judicial decision. In marked contrast, five of the six of the diagnostic testing cases settled before any substantive judicial decision, and the sixth was dismissed soon after

the lawsuit was filed because the plaintiff lacked sufficient ownership interest.

Litigations involving protein therapeutics also tend to remain active longer than other human gene patent litigations (*13*). The time from filing to resolution ranged from 23 to 112 months for therapeutic protein litigations, resulting in an average pendency of 63 months (*19*). Cases involving genetic diagnostic testing were all settled or dismissed within 2 to 17 months of filing, resulting in an average pendency of only 8 months.

Implications

Patents can have an impact, even when they are not asserted in court (*1*, *3*, *11*, *13*). However, for the most part, fears expressed concerning human gene patents have not been manifested overtly in patent litigation. Human gene patent litigation invariably has involved an alleged infringer engaged in substantial commercial activities focused specifically on the single gene that is the subject of the asserted patent, the antithesis of a patent thicket scenario (*14*). Some have speculated that DNA microarray technology is particularly at risk of becoming entangled in a thicket (*6*). However, I found no instance in which a human gene patent was asserted against the manufacturer or user of microarray technology, although microarray companies have experienced substantial patent litigation involving nongene patents since the mid-1990s.

Many gene patents only claim some limited use of a gene and, thus, do not preclude use of the gene in a different system or context (*14*). For example, one of the patents characterized by Jensen and Murray as "claiming a human gene" only covers the use of a mammalian cell culture system to express and secrete the product of an exogenous, recombinant human α-galactosidase A gene (*20*). Transkaryotic Therapies designed around the patent by expressing the endogenous gene in cultured human cells (*14*). Gene patents can also be avoided by taking research off shore. For example, in the research tool context, a contract research organization (CRO) based in the United States successfully avoided patents' claiming the use of certain human genes in cell-

based assays by performing the assays in Taiwan and importing the data back to the United States (*14*). Likewise, genetic diagnostic testing service could be taken off-shore if patents prove a substantial impediment to access in the United States.

In four cases, an academic research institution was sued for infringement, but invariably the institution was involved in some substantial commercial enterprise focused on the patented human genes. For example, Myriad Genetics asserted its BRCA1 patents against the University of Pennsylvania for providing commercial BRCA1 genetic testing services, in direct

Pending Human Gene Patent Litigations

Forensic testing
Diagnostic testing
Research tool
Therapeutic protein
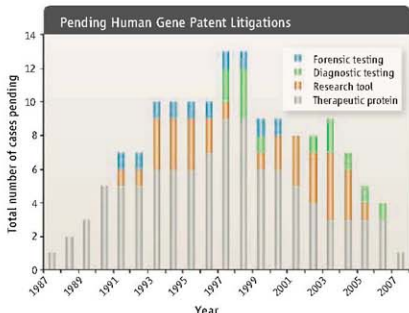
Year

Pending human gene patent litigations in each year starting in 1987 and extending to June of 2007. Two lawsuits resolved in the first part of 2007 are not included in the 2007 tally.

competition with Myriad (*14*). Yet, the human gene patents that have been asserted quite often arose out of academic research (*13*, *14*). For example, all but one of the human gene patent litigations that have been brought against a provider of genetic diagnostic testing services has involved an academic patent (*13*).

Most of the litigations have been brought in the context of therapeutic proteins, with the patents serving a function analogous to that of drug patents in the traditional pharmaceutical industry. Thus, these patents maintain companies' incentive to develop biotech drugs despite the expense and risk. My results provide little empirical support for a legislative bar to the patenting of genes or DNA.

If any legislative reform is deemed necessary, its scope would more appropriately be limited to some form of exemption from infringement liability for research and/or diagnostic testing uses of naturally occurring genetic sequences (*21*). Such an approach preserves the patentability of gene-based innovations, while curtailing some of the more problematic enforcement activities. Alternatively,

Congress could act to encourage funding agencies to exercise the march-in rights provided by the Bayh-Dole Act (*22*) in cases where a patent resulting from government funded research substantially impedes biomedical research or the availability of diagnostic testing. This approach was proposed by witnesses at a recent congressional hearing focused on gene patents (*23*).

References and Notes

1. T. Caulfield *et al.*, *Nat. Biotechnol.* **24**, 1091 (2006).
2. M. Crichton, *New York Times*, 13 February 2007, p. A23.
3. M. K. Cho *et al.*, *J. Mol. Diagn.* **5**, 3 (2003).
4. B. William-Jones, *Health Law J.* **10**, 123 (2002).
5. M. A. Heller, R. S. Eisenberg, *Science* **280**, 698 (1998).
6. J. H. Barton, *Nat. Biotechnol.* **24**, 939 (2006).
7. Genomic Research and Accessibility Act, H.R. 977, 110th Congr. (2007).
8. J. P. Kushan, on behalf of the Biotechnology Industry Organization (BIO), before the House Judiciary Subcommittee on Courts, the Internet, and Intellectual Property (30 October 2007); www.judiciary.house.gov/hearings/pdf/Kushan071030.pdf.
9. K. Jensen, F. Murray, *Science* **310**, 239 (2005).
10. *Cong. Rec.* **153**, E315 (daily ed., 9 February 2007).
11. J. P. Walsh *et al.*, *Science* **309**, 2002 (2005).
12. A "human gene patent" is any patent with a claim directed to a product or process that includes a single, specific human genetic sequence.
13. Methods and definitions are on *Science* Online and in (*14*).
14. An expanded discussion of the results of this survey can be found in C. M. Holman, *University of Missouri–Kansas City Law Rev.* **76**, 295 (2007); http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1090562.
15. M. A. Lemley, C. Shapiro, *J. Econ. Perspect.* **19**, 75 (2005).
16. Data obtained from the Stanford Intellectual Property Litigation Clearinghouse on 17 September 2008 for patents falling into U.S. Patent Class Classification Classes 514 and 424 (drugs, bio-affecting and body treating compositions) and 435 (chemistry: molecular biology and microbiology).
17. L. Pressman *et al.*, *Nat. Biotechnol.* **24**, 31 (2006).
18. M. M. Hopkins *et al.*, *Nat. Biotechnol.* **25**, 185 (2007).
19. Litigation pendency refers to the time between the filing of a lawsuit and resolution of the litigation.
20. U.S. Patent no. 5,356,804.
21. See, e.g., Genomic Research and Diagnostic Accessibility Act of 2002, H.R. 3967, 107th Congress (2002).
22. 35 U.S. Code 203(a) (2008).
23. Hearing on "Stifling or Stimulating—The Role of Gene Patents in Research and Genetic Testing," before the Subcommittee on Courts, the Internet, and Intellectual Property, 110th Congress (30 October 2007); www.judiciary.house.gov/hearings/hear_103007.html.
24. Early settlement or dismissal—the parties either agreed to settle the case before any substantive judicial decision or the case was dismissed by the court; Late settlement—the parties agreed to settle the case subsequent to one or more substantive judicial decisions, but before appealing the decision to a higher court; Case appealed—the district court's decision was appealed to a higher court.
25. I thank L. Andrews, J. Carbone, R. Feldman, T. Holbrook, F. S. Kieff, A. Lara, N. Levit, C. Mason, G. Pulsinelli, M. Lemley, R. Cook-Deegan, A. Rai, K. Jensen and participants of the 2007 Intellectual Property Scholars Conference (9 to 10 August, Chicago, IL) for commentary, and K. Jensen and F. Murray for sharing their database of human gene patents.

Supporting Online Material
www.sciencemag.org/cgi/content/full/322/5899/198/DC1

10.1126/science.1160687

PALEONTOLOGY

Sauropod Gigantism

How did sauropod dinosaurs reach body sizes that remain unsurpassed in land-living animals?

P. Martin Sander[1] and Marcus Clauss[2]

Sauropod dinosaurs were the largest animals ever to inhabit the land (see the figure). At estimated maximum body masses of 50 to 80 metric tons, they surpassed the largest terrestrial mammals and nonsauropod dinosaurs by an order of magnitude. With body lengths of more than 40 m and heights of more than 17 m, their linear dimensions also remain unique in the animal kingdom. From their beginnings in the Late Triassic (about 210 million years ago), sauropods diversified into about 120 known genera. They dominated ecosystems for more than 100 million years from the Middle Jurassic to the end of the Cretaceous, setting a record that mammalian herbivores will only match if they can

[1]Division of Paleontology, Steinmann Institute, University of Bonn, D-53115 Bonn, Germany. [2]Clinic for Zoo Animals, Exotic Pets and Wildlife, Vetsuisse Faculty, University of Zurich, CH-8057 Zurich, Switzerland. E-mail: martin.sander@uni-bonn.de; mclauss@vetclinics.uzh.ch

double their current geological survival time. Thus, sauropods were not only gigantic but also, in evolutionary terms, very successful. Recent advances bring us closer to understanding the enigma of their gigantism (1–3).

Extrinsic causes have repeatedly been advanced to explain the success of sauropod dinosaurs and the gigantism seen in the dinosaur era. However, physical and chemical conditions in the Mesozoic (250 to 65 million years ago) were probably less favorable for plant and animal life than they are today; for example, atmospheric O_2 concentrations were much lower (4). The variation of other factors (such as landmass size, ambient temperature, and atmospheric CO_2 concentrations) through time is not tracked by variations in sauropod body size (2, 5). Thus, the clue to sauropod gigantism must lie in their unusual biology (see the figure).

Sauropods had an elephantine body supported by four columnar legs and ending in a long tail. From the body arose a long neck bearing a small skull. Sauropods exhibit diverse oral, dental, and neck designs, indicating dietary niche differentiation; this variety makes reliance on any particular food source (6) as the reason for gigantism unlikely. However, one evolutionarily primitive character truly sets sauropods apart: In contrast to mammals and advanced bird-hipped dinosaurs (duck-billed and horned dinosaurs), they did not masticate their food; nor did they grind it in a gastric mill, as did some other herbivorous dinosaurs (7). Because gut capacity increases with body mass (8), the enormous gut capacity of sauropods would have guaranteed the long digestion times (6) necessary for degrading unchewed plant parts, even at a relatively high food intake.

The lack of a masticatory apparatus allowed sauropod heads to remain small and was one prerequisite for their long neck to
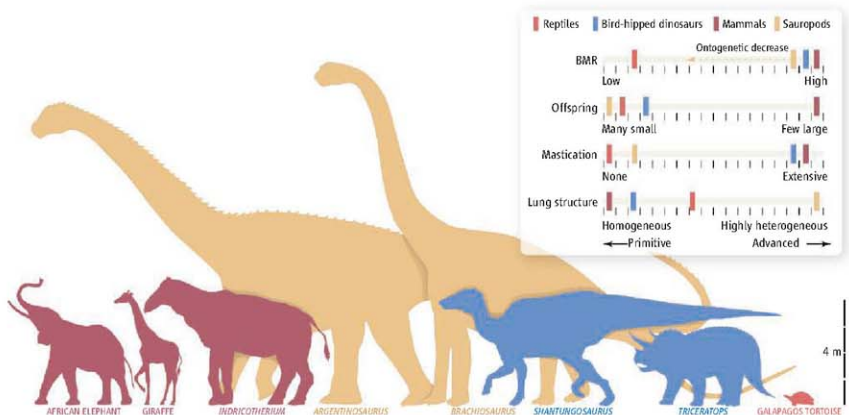
Toward understanding sauropod dinosaur gigantism. The sauropod dinosaurs *Brachiosaurus* and *Argentinosaurus* were much larger than the largest bird-hipped dinosaurs *Shantungosaurus* and *Triceratops*, the fossil rhinoceros *Indricotherium* (the largest known land mammal), the African elephant, the giraffe, and the Galapagos tortoise (the largest living herbivorous reptile). **(Inset)** The main biological properties that control the upper limits of body size in terrestrial herbivores—sauropod dinosaurs, bird-hipped dinosaurs, mammals, and ectothermic herbivorous reptiles—are visualized as sliders, with the evolutionarily primitive state to the left and the advanced state to the right. The slider position for each herbivore group (color-coded to match the images) indicates the specific combination of primitive and advanced states that led to the maximal body size of this group. The unique gigantism of sauropod dinosaurs was made possible by a high basal metabolic rate (BMR, advanced), many small offspring (primitive), no mastication (primitive), and a highly heterogeneous lung (advanced). We hypothesize that ontogenetic flexibility of BMR was also important.

CREDIT: ADAPTED BY P. HUEY/SCIENCE

evolve. This neck would have provided access to food out of reach of other animals (9) with little foraging movement of the whole body, while probably also functioning as a display organ and hence being subject to runaway sexual selection (10). In herbivorous mammals and evolutionarily advanced bird-hipped dinosaurs, in contrast, a disproportionate evolutionary increase in the size of the head and the masticatory apparatus relative to the rest of the body set a mechanical limit to the evolution of very long necks. In addition, the vulnerability of a long neck (10) makes its evolution unlikely without protection from predation by superior body size.

The long neck and large body size pose particular physiological problems. Large body size in endothermic animals is associated with a major problem of dissipating excess body heat. A long neck also means that a large volume of air must be moved in the windpipe during ventilation before fresh air reaches the lung. These problems appear to have been solved by an evolutionary innovation shared by sauropods and theropods (meat-eating dinosaurs) and their descendants, the birds: a highly heterogeneous avian-style respiratory system (11) with cross-current gas exchange in the lung and air sacs that pneumatized the vertebrae of the neck and the trunk and filled large parts of the body cavity. Compared to mammalian or reptilian lungs, this system overcame the problem of the long windpipe of sauropods (11) and also probably helped to dissipate excess body heat via the visceral air sac surfaces (11, 12).

For selective advantages conferred by large body size to be effective (13), this large body size must be reached quickly by the individual. Uniquely among amniotes, sauropods grew through five orders of magnitude from a 10-kg hatchling to a 100,000-kg fully grown individual. Bone histological evidence indicates that this growth took place at rates comparable to those of large terrestrial mammals (14,15); reproductive maturity was reached in the second decade of life and full size in the third decade of life (15), as predicted by demographic models that show higher ages at first reproduction to be incompatible with long-term population persistence (16). Such high growth rates are seen only in animals with a basal metabolic rate (BMR) of mammals and birds (17). For sauropods, rather than assuming a constant metabolic rate throughout the animals' life, an ontogenetic decrease in BMR has been suggested (12, 18). Such a decrease would reconcile rapid growth rates in juveniles with problems resulting from gigantic body size (such as overheating and high food requirements) in adults.

Considering the costs of a high BMR, it may have evolved early on in sauropods, as an adaptation for the high growth rates necessary for reaching very large body size.

At the population level, egg-laying and the associated production of many small offspring (19), in contrast to the one-offspring strategy of mammalian megaherbivores, is a key characteristic of sauropod reproductive biology and of the dinosaur ecosystem (20, 21). This strategy may have guaranteed long-term survival of gigantic species (20). In mammals, large body size increases the risk of chance extinction by reducing population density and increasing population recovery time: With increasing body size, fewer offspring are produced, and these take longer to mature. The retention of the primitive feature of egg-laying might have alleviated this constraint through much higher population recovery rates than in large mammals (21).

Thus, we suggest that the unique gigantism of sauropods was made possible by a combination of phylogenetic heritage (lack of mastication, egg-laying) and a cascade of evolutionary innovations (high growth rate, avian-style respiratory system, and a flexible metabolic rate). Although modern mammals evolved a high growth rate independently, the comparison with sauropods identifies the mastication of food (inherited from small insectivorous ancestors), overheating caused by an inadequate cooling mechanism, and giving birth to live offspring as the major factors limiting the potential body size of mammalian herbivores (see the figure).

Recent modeling studies found a high metabolic rate to be incompatible with gigantic body size because of the problem of heat dissipation (22, 23); however, these models assumed metabolic rate to have remained constant throughout life. In addition, one of the studies (23) suffers from poor bone histologic constraints on sauropod growth rates, as does a study (24) arguing against fast growth in sauropods. Compared to other dinosaurs, the long bones of sauropods rarely preserve growth marks, probably because bone tissue was deposited too rapidly to record them (15). Histologic growth rate studies using skeletal elements other than long bones may provide more reliable estimates.

Theropod dinosaurs such as *Tyrannosaurus rex* are similarly outsized compared to mammalian carnivores (1) as sauropods are compared to mammalian herbivores. Thus, there may be links between sauropod gigantism and meat-eating dinosaur gigantism. One such link may be the mode of reproduction. Egg-laying of sauropods must have made large amounts of food available to predators in

the form of many small, little-protected young sauropods. In contrast, mammalian megaherbivores withhold this food source from carnivores by rearing very few, well-protected offspring. This greatly decreased resource base may limit maximum body size of carnivores today (21).

Further progress in understanding dinosaur gigantism will come from a conservation biology approach that models the carrying capacity of Mesozoic ecosystems based on the juvenile-biased population structure of dinosaurs (19, 20). Such an approach, which goes beyond the reconstruction of an individual's metabolism (14, 15, 22, 23), will be much more informative regarding resource limitations, population growth potential, and body size evolution of dinosaurs.

References and Notes
1. G. P. Burness, J. Diamond, T. Flannery, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 14518 (2001).
2. M. T. Carrano, in *Amniote Paleobiology: Perspectives on the Evolution of Mammals, Birds, and Reptiles*, M. T. Carrano, R. W. Blob, T. J. Gaudin, J. R. Wible, Eds. (Univ. of Chicago Press, Chicago, 2006), pp. 225–268.
3. P. M. Sander, O. Mateus, T. Laven, N. Knötschke, *Nature* **441**, 739 (2006).
4. R. Berner, J. M. VandenBrooks, P. D. Ward, *Science* **316**, 557 (2007).
5. D. W. E. Hone, T. M. Keesey, D. Pisani, A. Purvis, *J. Evol. Biol.* **18**, 587 (2005).
6. J. Hummel *et al.*, *Proc. R. Soc. B* **275**, 1015 (2008).
7. O. Wings, P. M. Sander, *Proc. R. Soc. B* **274**, 635 (2007).
8. M. Clauss, A. Schwarm, S. Ortmann, W. J. Streich, J. Hummel, *Comp. Biochem. Physiol. A* **148**, 249 (2007).
9. A. Christian, G. Dzemski, *Fossil Rec.* **10**, 38 (2007).
10. P. Senter, *J. Zool.* **271**, 45 (2007).
11. M. J. Wedel, *Paleobiology* **29**, 243 (2003).
12. S. F. Perry, A. Christian, T. Breuer, N. Pajor, J. R. Codd, *J. Exp. Zool. A*.
13. L. N. Cooper, A. H. Lee, M. L. Taper, J. R. Horner, *Proc. R. Soc. B*, 10.1098/rspb.2008.0912 (2008).
14. G. M. Erickson, K. C. Rogers, S. A. Yerby, *Nature* **412**, 429 (2001).
15. P. M. Sander, *Paleobiology* **26**, 466 (2000).
16. A. E. Dunham, K. L. Overall, W. P. Porter, C. A. Forster, in *Paleobiology of the Dinosaurs*, J. O. Farlow, Ed. (Geological Society of America, Boulder, CO, 1989), pp. 1–21.
17. T. J. Case, *Q. Rev. Biol.* **53**, 243 (1978).
18. J. O. Farlow, in *The Dinosauria. 1st Edition*, D. B. Weishampel, P. Dodson, H. Osmolska, Eds. (Univ. of California Press, Berkeley, CA, 1990), pp. 43–55.
19. P. M. Sander, C. Peitz, F. D. Jackson, L. M. Chiappe, *Palaeontographica A* **284**, 69 (2008).
20. C. M. Janis, M. Carrano, *Acta Zool. Fenn.* **28**, 201 (1992).
21. J. Hummel, M. Clauss, *Evol. Ecol. Res.* **10**, 925 (2008).
22. F. Seebacher, *Paleobiology* **28**, 105 (2003).
23. J. F. Gillooly, A. P. Allen, E. L. Charnov, *PLoS Biol.* **4**, 1467 (2006).
24. T. M. Lehman, H. N. Woodward, *Paleobiology* **34**, 264 (2008).
25. The research of P.M.S. and M.C. was supported by the German Research Foundation (grant FOR 533). This is contribution number 46 of the DFG research unit 533 "Biology of the Sauropod Dinosaurs: The Evolution of Gigantism."

10.1126/science.1160904

IMMUNOLOGY

Regulating Suppression

Ethan M. Shevach

How a T cell protein suppresses the immune response will help guide the development of therapies that do not have autoimmune side effects.

The vertebrate immune system has important self-check mechanisms that prevent it from destroying the host's own tissues. Regulatory T cells are part of this suppressive control, and on page 271 in this issue, Wing *et al.* (*1*) define the role of cytotoxic T lymphocyte antigen 4 (CTLA-4)—a protein that is preferentially expressed by regulatory T cells—in restraining the immune response when needed and in maintaining self tolerance. Therapies based on CTLA-4 are currently used, or in clinical trials, for treating autoimmune conditions such as rheumatoid arthritis or to augment immune responses directed at tumors, respectively, so understanding the specific role of CTLA-4 in regulatory T cell function may further improve and expand such clinical approaches.

Optimal T cell proliferation and acquisition of effector functions require intracellular signals elicited by both the T cell receptor and by a co-receptor, CD28 (*2*). Although CTLA-4 binds to the same ligands as CD28 (with much higher affinity), it restricts T cell activation. CD28 and CTLA-4 also differ in their expression on different subsets of T cells. CD28 is expressed on most CD4+ T cells and almost all CD8+ T cells. CTLA-4 expression is induced in all T cells transiently after T cell receptor activation. However, among non-activated (resting) CD4+ T cells, CTLA-4 is constitutively expressed on regulatory T cells that express the forkhead transcription factor, Foxp3 (*3*).

To clearly define a role for CTLA-4 in regulatory T cell function, Wing *et al.* deleted expression of the protein in mice, specifically in regulatory T cells that express Foxp3. These animals appeared to be normal until 7 weeks of age, when they rapidly developed systemic autoimmunity. Thus, CTLA-4 deficiency in regulatory T cells alone is sufficient to cause fatal disease, and maintenance of its expression in activated effector T cells is insufficient to prevent this outcome.

Previous studies suggested that CTLA-4 might be involved in mediating the suppressor function of Foxp3-expressing regulatory T cells both in vivo (*4*, *5*) and in vitro (*6*, *7*). But interpretation of these studies is problematic,
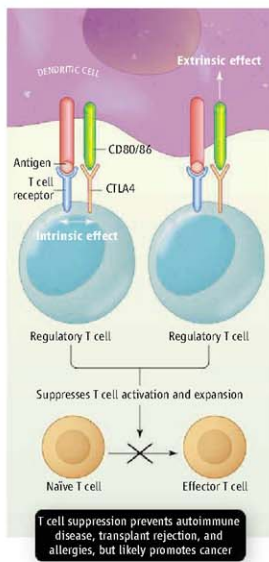
Signaling suppression. A possible mechanism of suppressing T cell activity involves engagement of CTLA-4 on regulatory T cells with CD80/CD86 on dendritic cells, with the resultant delivery of an extrinsic signal that limits dendritic cell maturation. An intrinsic signal elicited by CTLA-4 function in regulatory T cells has not been ruled out.

as they used treatment with an antibody to CTLA-4, or regulatory T cells from mice with a global deficiency of CTLA-4. Anti–CTLA-4 may have abrogated suppression by indirectly activating effector T cells. Although mice lacking CTLA-4 do have Foxp3-expressing regulatory T cells, isolating them from these animals has required immunologic manipulations that may have influenced regulatory T cell development (*5*, *7*). These problems are absent in the analysis of regulatory T cells from the mice of Wing *et al.* that have a selective loss of CTLA-4. Selective CLTA-4 deficiency did not appear to alter the develop-

ment or homeostasis of regulatory T cells, nor render them pathogenic. These cells also remained anergic but were less suppressive in vitro than their wild-type counterparts. Most notably, regulatory T cells from these mice were less suppressive in vivo, as immunodeficient mice reconstituted with total CD4+ T cells from these mice had enhanced immune responses to transplanted tumors.

Although these experiments support the view that the expression of CTLA-4 is required for optimal function of Foxp3-expressing regulatory T cells, the mechanism of action of this protein remains unclear (see the figure). Indeed, considerable controversy also exists with regard to the mechanisms by which CTLA-4 leads to inhibiting the activation of effector T cells. Most studies (*8–10*) support the view that CTLA-4 expressed on activated effector T cells, by a ligand-dependent or even in a ligand-independent pathway, delivers a signal into the T cell that inhibits its activation at the level of the T cell receptor.

On the other hand, rather than such an intrinsic signal, Wing *et al.* propose an extrinsic signaling mechanism based on their observations and those of others (*11*, *12*) that the interaction of CTLA-4 on regulatory T cells with CD80 and CD86, ligands on dendritic cells (immune cells that present antigen to T cells), blocks the subsequent increase of CD80/CD86 expression (or even down-regulates CD80/CD86 expression) induced by antigen-specific effector T cells. Regulatory T cells from the mice lacking CTLA-4 were defective when compared to regulatory T cells from wild-type mice in preventing the increased expression of CD80 and CD86 in dendritic cells (*3*). The addition of an anti–CTLA-4 fragment into cultures of wild-type regulatory T cells and dendritic cells partially inhibited the suppression of CD80/ CD86 expression (*11*). Presumably, inhibition of CD80/CD86 expression by regulatory T cells limits the capacity of the dendritic cells to stimulate naïve T cells through CD28, resulting in immune suppression. The biochemical nature of the extrinsic signal transduced to a dendritic cell by the interaction of CTLA-4 with CD80/CD86 remains to be characterized. An intrinsic function of CTLA-4 in the regulatory T cell cannot be excluded and deserves further investigation.

Polymorphisms in CTLA-4 have been implicated in the susceptibility to multiple

Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, Bethesda, MD 20892, USA. E-mail: eshevach@niaid.nih.gov

CREDIT: P. HUEY/SCIENCE

autoimmune diseases, including diabetes and multiple sclerosis (*13*). Do these polymorphisms directly affect regulatory T cell function? A fusion protein composed of CTLA-4 and immunoglobulin protein (called Abatacept, developed by Bristol-Meyers-Squibb) has been approved to treat rheumatoid arthritis and is thought to inhibit stimulation of T cells by blocking the interaction of CD80/CD86 with CD28 (*14*). Could this therapeutic regimen also inhibit regulatory T cell function and promote an autoimmune response after prolonged administration? Anti–CTLA-4 is now in clinical trials to boost tumor immunity and is thought to stimulate antitumor effector T cells by blocking inhibitory signaling by CTLA-4 (*15*). As the development of autoimmune disease is a side effect of this therapy, it remains to be determined whether some of the effects of the CTLA-4–specific antibody are secondary to an inhibition of regulatory T cell function that may be required for an optimal anti-tumor response.

References

1. K. Wing *et al.*, *Science* **322**, 271 (2008).
2. B. M. Carreno, M. Collins, *Annu. Rev. Immunol.* **20**, 29 (2002).
3. S. F. Ziegler, *Annu. Rev. Immunol.* **24**, 209 (2006).
4. S. Read, V. Malmstrom, F. Powrie, *J. Exp. Med.* **192**, 295 (2000).
5. S. Read *et al.*, *J. Immunol.* **177**, 4376, (2006).
6. T. Takahashi *et al.*, *J. Exp. Med.* **192**, 303 (2000).
7. Q. Tang *et al.*, *Eur. J. Immunol.* **34**, 2996 (2004).
8. B. T. Fife, J. A. Bluestone, *Immunol. Rev.* **212**, 166 (2006).
9. D. M. Sansom, L. S. K. Walker, *Immunol. Rev.* **212**, 131 (2006).
10. T. Pentcheva-Hoang *et al.*, *Immunity* **21**, 401 (2004).
11. Y. Onishi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 10113 (2008).
12. C. Oderup *et al.*, *Immunology* **118**, 240 (2006).
13. H. Ueda *et al.*, *Nature* **423**, 506 (2003).
14. J. M. Kremer *et al.*, *N. Engl. J. Med.* **349**, 1907 (2003).
15. G. Q. Phan *et al.*, *Proc. Natl. Acad. Sci.* **100**, 8372 (2003).

PHYSICS

Cold Molecules Beat the Shakes

P. L. Gould

Due to their relative complexity, molecules have been harder to cool than atoms, but that is beginning to change.

Laser cooling of atoms began more than 25 years ago and has led to spectacular advances in areas such as atomic clocks, Bose-Einstein condensation, degenerate Fermi gases, and ultracold collisions. Molecules, due to their complexity relative to atoms (*1*), have been slower to jump on the ultracold bandwagon but are now firmly on board. Three recent papers (*2–4*) describe techniques for efficiently and substantially reducing the vibrational motion of molecular ensembles, thus approaching the goal of large numbers of molecules that are cold in all their degrees of freedom, both external and internal.

In molecules, not only can electrons be excited but also the atoms themselves can vibrate and rotate, leading to a multitude of energy levels. Traditional laser cooling relies on momentum kicks from repeated cycles of photon absorption and emission and therefore only works if the excited state decays back to the original ground state. Certain states, such as alkali metals, have these "cycling" transitions; molecules do not. Alternative methods for producing cold molecules are therefore required. Along the way, an important question arises: What is meant by molecular temperature? With atoms, there is little ambiguity: Temperature describes their translational motion. However, molecules also have internal energy due to vibration and rotation. Molecules can be very cold translationally but still have high vibrational energy.

Cold molecule production has generally followed one of two strategies: Either start with hot molecules and cool them, or start with cold atoms and assemble them into molecules. Examples of the first category include cooling by collisions with cold He buffer gas (*5*) and slowing a molecular beam with pulsed electric fields (*6*). These direct cooling techniques can provide molecules in their lowest internal state, but the sub-kelvin translational temperatures achieved so far are not considered ultracold. The second category includes photoassociation (*7*), in which laser light excites a pair of cold atoms into a bound molecular state, and the process of magneto-association (*8*), whereby a ramped magnetic field is applied to bind the atoms. Both techniques yield extremely low translational temperatures (sub-mK or sub-μK) but generally leave the molecules in barely bound levels of high vibrational excitation. Although there has been some progress in populating the lowest vibrational states (*9–11*), the recent breakthrough (*2–4*) is true vibrational cooling: the efficient removal of a large fraction of this vibrational energy.

Viteau *et al.* (*2*) photo-associate ultracold Cs atoms to form Cs_2 molecules in the lowest electronic state but in a distribution of vibrational levels. By repeatedly optically pumping these molecules

with broadband light to an electronically excited state (see the figure), which then radiatively decays back to the ground electronic state, the excited vibrational levels are eventually transferred into the lowest possible vibrational level, $v = 0$. The optical pumping
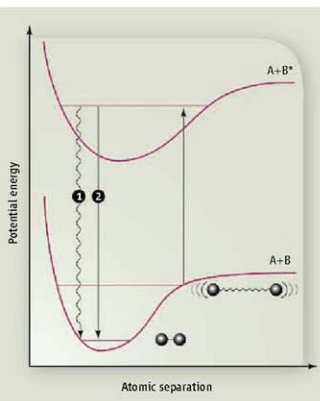
Cool it. Potential energy curves of a diatomic molecule AB. The lower curve is the ground electronic state; the upper curve is an excited electronic state. Ultracold molecules are produced initially in the ground electronic state but in a level of high vibration. They are transferred to a lower level of vibration by absorbing laser light (upward arrow), then returning to the ground state by either radiative decay [path 1, used in (*2*)] or stimulated emission with another laser [path 2, used in (*3, 4*)].

Department of Physics, University of Connecticut, Storrs, CT 06269–3046, USA. E-mail: phillip.gould@uconn.edu

is special in two regards. First, it uses ultrafast laser pulses that are spectrally broad enough to excite all the relevant vibrational levels. Second, the spectrum is shaped to eliminate all frequencies that would excite from $v = 0$. As a result, molecules are trapped in $v = 0$, so after many incoherent cycles of excitation and decay, they accumulate in this "dark" state. Ironically, it is exactly this dark-state population trapping that has prevented traditional laser cooling of molecular motion. Here, it is put to good use: vibrationally cooling molecules that are already translationally cold.

Danzl *et al.* (3) also work with Cs_2 but use magneto-association of a Bose-Einstein condensate of Cs atoms to initially form the molecules. Rather than relying on broadband light to induce multiple absorption/emission cycles, they use a pair of laser beams with precisely defined frequencies to coherently drive the population from the initial state of high vibration, through an electronically excited intermediate state, then back down to a state of low vibration (see the figure). One difficulty with this process is the poor overlap of the wave functions of the highly excited and low-lying vibrational states. However, the lasers used for this two-photon process are locked to a frequency comb, and therefore highly coherent, allowing a long interaction time (10 μs) and efficient transfer (80%) to the lower energy state. Danzl *et al.* can remove 0.13 eV of vibrational energy, which puts them one-fourth of the way to $v = 0$. They are optimistic

that by applying one more judiciously chosen two-photon process, they can reach the absolute ground state.

The technique used by Ni *et al.* on page 231 of this issue (4) is very similar to that of Danzl *et al.* but is applied to a rather different molecule, KRb. They report 56% transfer from the barely bound initial state to $v = 0$ of the lowest triplet electronic state, which is bound by 0.03 eV. Even more impressive is their demonstration of 83% transfer, using an intermediate state of mixed triplet-singlet character, to $v = 0$ of the lowest singlet electronic state. This is the absolute ground state of the system, bound by a whopping 0.52 eV. The fact that the KRb molecule is composed of two different atoms means that, as observed in this experiment, it possesses an electric dipole moment.

There is currently a great deal of interest in dipolar systems at low temperatures and high densities. Interactions between dipoles are both long range and anisotropic: Two dipoles oriented head-to-tail attract; side-by-side they repel. So, for example, a confined pancake-shaped sample will tend to be stable, whereas a cigar-shaped sample will tend to collapse. Such dipolar effects have begun to be observed in systems of magnetic dipoles (12), but the interactions will be much stronger between electric dipoles, enabling applications such as the modeling of complex many-body systems. Dipole-dipole interactions may also enable communication between cold molecule qubits in a quantum computer (13) and affect ultracold chemi-

cal reactions. For all these potential applications, a large dipole moment is desired. In states of high vibrational excitation, the atoms live far apart and the dipole moment is small; hence, the motivation for eliminating vibration.

Another compelling reason for going to the absolute ground state is a purely practical one: stability. All other states are unstable against inelastic collisions, which is a problem at high density. This recent progress toward populating the lowest-energy state therefore bodes well for producing a stable Bose-Einstein condensate or degenerate Fermi gas of molecules. Such systems will prove useful in exploring exotic quantum phases of matter and performing quantum simulations of highly correlated condensed matter systems.

References

1. J. T. Bahns, W. C. Stwalley, P. L. Gould, *J. Chem. Phys.* **104**, 9689 (1996).
2. M. Viteau *et al.*, *Science* **321**, 232 (2008).
3. J. G. Danzl *et al.*, *Science* **321**, 1062 (2008).
4. K.-K. Ni *et al.*, *Science* **322**, 231 (2008).
5. J. D. Weinstein *et al.*, *Nature* **395**, 148 (1998).
6. H. L. Bethlem, G. Berden, G. Meijer, *Phys. Rev. Lett.* **83**, 1558 (1999).
7. K. M. Jones, E. Tiesinga, P. D. Lett, P. S. Julienne, *Rev. Mod. Phys.* **78**, 483 (2006).
8. T. Köhler, K. Góral, P. S. Julienne, *Rev. Mod. Phys.* **78**, 1311 (2006).
9. A. N. Nikolov *et al.*, *Phys. Rev. Lett.* **84**, 246 (2000).
10. J. M. Sage, S. Sainis, T. Bergeman, D. DeMille, *Phys. Rev. Lett.* **94**, 203001 (2005).
11. J. Deiglmayr *et al.*, arXiv: 0807.3272 (2008).
12. T. Lahaye *et al.*, *Nature* **448**, 672 (2007).
13. D. DeMille, *Phys. Rev. Lett.* **88**, 067901 (2002).

EVOLUTION

Armor Development and Fitness

William A. Cresko

The fitness of stickleback fish that develop different numbers of external bony plates varies between oceanic and freshwater environments.

Nearly 150 years after the publication of the *Origin of Species*, it is humbling to contemplate how well Darwin outlined the processes of evolution. The heritable basis of traits confounded him, however, and evolutionary biologists have since attempted to connect the processes of natural selection and genetic drift with the origin and distribution of genetic variation in the wild (1, 2). A flurry of recent work mapping phenotype to genotype has identified the molecular genetic basis of some traits in natural populations (3, 4), but documenting the fitness consequences of these genes has been more elu-

sive. An important study by Barrett *et al.* on page 255 in this issue (5) attempts to fill this gap by studying changes in allele frequencies in replicate populations of the threespine stickleback (*Gasterosteus aculeatus*), thereby adding an intriguing new wrinkle to a rapidly developing story.

Stickleback that originate in the ocean and subsequently become isolated in freshwater environments rapidly evolve the loss of external bony lateral plates (6). A major genetic factor on linkage group IV (the part of the genome corresponding to the largest stickleback chromosome) was implicated in this loss of armor in multiple populations (7, 8). "Low" alleles of the gene *ectodysplasin-A* (*Eda*)— which encodes a signaling molecule involved in ectodermal outgrowths such as hair, scales,

and teeth—were subsequently associated with the loss (9). Barrett *et al.* screened thousands of oceanic stickleback to find those heterozygous at the *Eda* locus—carrying one low and one "complete" allele, the latter of which encodes for the development of a full set of lateral plates. These fish were then introduced into freshwater ponds and quickly produced offspring that were sampled and genotyped at *Eda* throughout a full year (one stickleback generation). As expected, there was very strong selection (a selection coefficient of $s \sim 0.5$) for the low *Eda* allele in offspring fish that became large enough to have developed the complete set of plates (see the figure). This fitness differential could be due to the burden of forming and maintaining lateral plates, with the low *Eda* allele conferring higher

Center for Ecology and Evolutionary Biology, Department of Biology, University of Oregon, Eugene, OR 97403–5289, USA. E-mail: wcresko@uoregon.edu
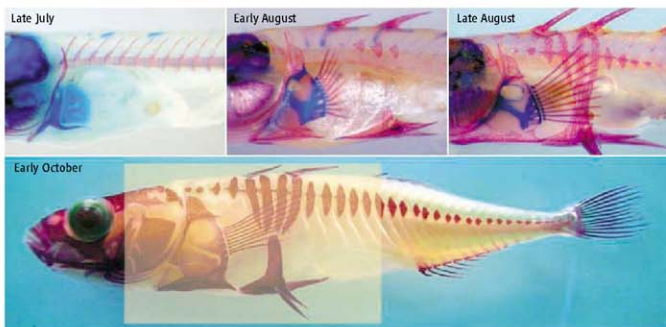
Stickleback lateral plates. Approximate development of lateral plates in the threespine stickleback is shown for a fish born in early July: the absence of plates in late July; the formation of anterior lateral plates in early August; the formation of posterior lateral plates in late August; and the full complement of plates in early October. These stickleback are from a line grown in the laboratory, originally derived from a wild Alaskan oceanic population, and are used for illustrative purposes. The graph represents data from Barrett *et al.* showing selection against ($s < 0$), and for ($s > 0$), the low *Eda* allele and the heterozygous genotype, before and after fish develop a full complement of plates, respectively.
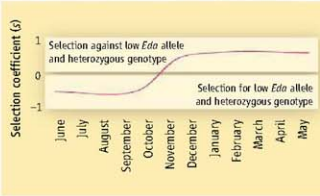
fitness in fresh waters because energy can be shunted away from bone formation and toward growth and reproduction (*10*). This "cost of plates" argument has intuitive appeal because freshwater stickleback populations at high latitudes need to endure many winter months, and increasing growth rate and sequestering energy could increase the probability of surviving the winter and reproducing early in the spring.

Despite the strong selection for the low *Eda* allele in older (larger) fish, selection was actually against this same allele before and during plate formation in young (smaller) offspring. Importantly, the selection coefficients against, and then for, the low *Eda* allele in young and older fish, respectively, were almost equal, a curious result considering the very rapid and nearly universal loss of plates once stickleback invade fresh waters (*11*). Barrett *et al.* hypothesize that *Eda* alleles may influence other phenotypic traits that are important for fitness before the plates fully develop. Alternatively, the change in *Eda* allele frequency in young fish may be a correlated response to selection of alleles of linked loci.

Freshwater habitats differ from oceanic ones in many conditions to which invading stickleback must adapt. Heterozygous fish were collected in the wild, and not created

through allelic introgression (the experimental addition of only the low *Eda* allele). Therefore, the genotype of the *Eda* allele may actually be a marker for haplotypes (genes closely associated on a chromosome that are inherited together) that contain *Eda* and numerous other genes in this region of linkage group IV. The linkage in a single haplotype of alleles at different loci subject to opposing selection could subsequently be broken through gene recombination during reproduction in the freshwater environment. Fixation of the recombinant low *Eda* allele–containing haplotype could occur due to the combined effects of selection for the low *Eda* allele and for a linked allele, producing the commonly observed rapid lateral plate evolution. Depending on the physical distance between linked genes on a chromosome and the recombination rate (the rate of exchange of DNA sequence between homologous chromosomes, thus forming new combinations of alleles), this pattern might be observed only after multiple generations. The linkage hypothesis makes the testable prediction that although many low *Eda* alleles may be present when a freshwater population is founded, a "hard" selective sweep for this region would involve only one or few recombinant haplotypes in any particular population (*12*).

Surprisingly, changes in the frequency of the *Eda* allele are mediated largely through selection on heterozygotes that varies as the fish mature, implying a role for complex modifications to *Eda* genetic interactions—both dominance (between *Eda* alleles) and epistasis (between *Eda* and other genes)—during the formation of lateral plates. These data are consistent with recent work showing patterns of long-distance (>25 megabases) linkage disequilibrium (nonrandom association of genomic regions) across linkage group IV in laboratory crosses in multiple stickleback populations (*13*). Together, these data hint that chromosomal structures such as inversions and translocations of DNA fragments from one location to another, as well as epistatic interactions among linked loci, may play an important role in the evolution of this genomic region and of the lateral plate phenotype. Amazingly, most of the present *Eda* low alleles that exist in stickleback populations around the world appear to be derived from a single allele that originated around 2 million years ago (*9*), and therefore most low alleles associated with lateral plate loss are likely identical by descent. Thus, the *Eda* alleles, and presumably other regions of the stickleback genome, have interacted in alternative oceanic and freshwater environments over very long periods, with potentially important consequences for the organization of genetic variation in the genome. A challenge for geneticists is to understand how evolutionary processes affect genetic variation in spatially explicit ways, both across geography and within genomes (*14*).

Despite these additional questions, the study by Barrett *et al.* is an excellent example of how experimental approaches can be used to link alleles to phenotypes and ultimately to fitness. Understanding the mechanistic basis of identified alleles in terms of cell behaviors, developmental trajectories, and physiological systems (*15*) may help to explain how intermediaries in the genotype-to-phenotype map might influence the distribution of phenotypic variation in new environments or lead to pervasive parallel evolution such as that seen

in stickleback. The study by Barrett *et al.* points to the great promise for connecting molecular genetics with phenotypic variation and fitness in the wild, a synthesis that would have made a pleasant gift for Darwin on his 200th birthday next February.

References and Notes
1. R. A. Fisher, *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1930).
2. S. Wright, *Evolution and the Genetics of Populations* (University of Chicago Press, Chicago, 1984), vols. 1 to 4.
3. S. B. Carroll, *Cell* **134**, 25 (2008).
4. D. L. Stern, V. Orgogozo, *Evolution* **62**, 2155 (2008).
5. R. D. H. Barrett, S. M. Rogers, D. Schluter, *Science* **322**, 255 (2008); published online 28 August 2008 (10.1126/science.1159978).
6. M. A. Bell, S. A. Foster, *The Evolutionary Biology of the Threespine Stickleback* (Oxford Univ. Press, Oxford, 1994).
7. P. F. Colosimo *et al.*, *PLoS Biol.* **2**, E109 (2004).
8. W. A. Cresko *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 6050 (2004).
9. P. F. Colosimo *et al.*, *Science* **307**, 1928 (2005).
10. N. Giles, *J. Zool.* **199**, 535 (1983).
11. M. A. Bell, W. E. Aguirre, N. J. Buck, *Evolution* **58**, 814 (2004).
12. R. D. Barrett, D. Schluter, *Trends Ecol. Evol.* **23**, 38 (2008).
13. M. R. Miller *et al.*, *Genome Res.* **17**, 240 (2007).
14. H. Ellegren, B. C. Sheldon, *Nature* **452**, 169 (2008).
15. A. M. Dean, J. W. Thornton, *Nat. Rev. Genet.* **8**, 675 (2007).
16. I thank S. Bassham, J. Snodgrass, and C. Hulslander for comments. W.A.C. is supported by grants from the NSF (IOS0642264 and IOS0818738) and NIH (R24GM79486).

ECOLOGY

Biodiversity in a Warmer World

A new framework helps to understand how species ranges change under global warming.

Jens-Christian Svenning[1] and Richard Condit[2]

There is ample evidence that 20th-century warming has shifted ranges of temperate and arctic species, but on page 261 of this issue, Moritz *et al.* (*1*) provide an exceptionally thorough example: They take advantage of a well-documented study from a century ago (*2*) to demonstrate contractions and expansions of elevation range among small mammals in Yosemite National Park, California, USA. In contrast, there have been few attempts to even address the tropics' sensitivity to global climate change (*3*). Also in this issue (page 258), Colwell *et al.* (*4*) use a novel conceptual approach to analyze climate shifts in tropical ecosystems.

Colwell *et al.* explain that weak latitudinal temperature gradients in the tropics will make it difficult for species to track suitable climatic conditions by migrating through the lowlands; instead, short-distance upslope migration to cooler mountains is what we should expect. The authors note three ways in which global warming may cause extinction. First, the tropical lowlands may experience biotic attrition: Warming drives species out of the lowlands, but no source of species adapted to higher temperatures is available to compensate the losses. The two additional risks for tropical mountain species are range-shift gaps (where species' current altitudinal ranges do not overlap climatically suitable ranges of the future) and mountain-top extinction (where warming pushes climatically suitable conditions off mountain peaks). These latter risks are also relevant outside the

[1]University of Aarhus, Ny Munkegade 1540, 8000 Aarhus C, Denmark. E-mail: svenning@biology.au.dk [2]Global Forest Observatory Network, Smithsonian Tropical Research Institute, Unit 0948, APO AA 34002, USA. E-mail: conditr@gmail.com
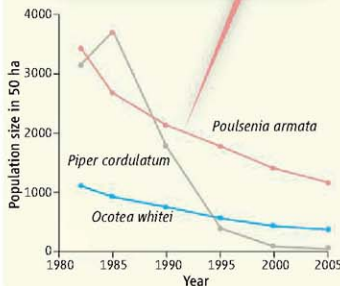
Steep decline. In a complete census of trees above 1 cm diameter in 50 ha of forest in Panama, the largest population declines are associated with drought, not temperature change (*13*).

tropics (*5*); indeed, Moritz *et al.* document the contraction of ranges of high-elevation species in Yosemite.

Colwell *et al.* then analyze ranges of 2000 species of plants and insects along a 2900-m altitudinal transect on Volcán Barva, Costa Rica, and relate these to expected upslope climate shifts. They find that a 3.2°C warming threatens 53% of the species with lowland extinction and 51% with range-shift gaps. Only a minority of species would face mountain-top extinction.

These numbers suggest large risks. However, the figures are likely to be controversial, because there are substantial uncertainties in our understanding of the sensitivity of tropical species to climatic warming. Notably, the prediction of heavy lowland extinctions is based on the assumption that species will be unable to tolerate temperatures higher than today's. Yet, many extant species evolved when climates were warmer (*6*) and may retain this warmth tolerance. Climatic limits within lineages often remain remarkably stable over millions of years (*7, 8*). On evolutionary time scales, there is little evidence that warming is detrimental in the tropics: Neotropical plant diversity peaked in the period of maximum warmth between 35 and 55 million years ago (*9*), and high tropical diver-

sity may be the product of the greater extent of warm areas in the past (10). On the other hand, extreme warming may have caused tropical extinctions or vegetation die-offs (11). Hence, the heat sensitivity of tropical lowland species is an open question.

Furthermore, Colwell et al. assume that temperature alone sets range limits. Although temperature is clearly a limiting factor for some tropical species at high altitudes, most studies on the distributions of lowland species focus on precipitation as limiting, because moisture has such an obvious impact (see the figure) (12, 13). As an example of moisture precedence, 30% of Panama's tree species limited to above 600 m above sea level on the dry Pacific slope occur near sea level on the wet Caribbean slope (14). Even where temperature is an important limiting factor, it is unlikely to be operating alone (15): The idiosyncratic range dynamics of small-mammal species at Yosemite (1) warn against the assumption that ranges simply reflect temperature tolerances. Finally, range limits estimated from small samples gathered at one location can

only be underestimates (16); predictions of extinction risk will be overestimated if ranges are underestimated.

Colwell et al. provide an important illustration of the potential risk posed by global warming to tropical diversity. In fact, even bleaker predictions have been made. One general circulation model predicts loss of Amazonian forest by the middle of this century due to drought stress (17). But forecasts of the impact of global warming on tropical diversity are hampered by uncertainties about what causes range limits. Even in temperate communities, little direct evidence of such factors goes into models; most models are based on correlations between current range and climate.

A key research focus should thus be to find direct evidence of how species respond to relevant environmental variables. The framework outlined by Colwell et al. can then be used more accurately, and will also be relevant outside the tropics. Even lowland attrition may occur here, because climatic shifts are likely to exceed species' migration capacities (18, 19).

References and Notes
1. C. Moritz et al., Science **322**, 261 (2008).
2. J. Grinnell, T. Storer, Animal Life in the Yosemite (Univ. California Press, Berkeley, CA, 1924).
3. J. A. Pounds et al., Nature **398**, 611 (1999).
4. R. K. Colwell, G. Brehm, C. L. Cardelús, A. C. Gilman, J. T. Longino, Science **322**, 258 (2008).
5. J. Lenoir et al., Science **320**, 1768 (2008).
6. E. Bermingham, C. Dick, Science **293**, 2214 (2001).
7. A. P. Moe, D. M. Smith, Palaeogeogr. Palaeoclimatol. Palaeoecol. **221**, 203 (2005).
8. V. Mosbrugger et al., Proc. Natl. Acad. Sci. U.S.A. **102**, 14964 (2005).
9. C. Jaramillo, M. J. Rueda, G. Mora, Science **311**, 1893 (2006).
10. J. J. Wiens, M. J. Donoghue, Trends Ecol. Evol. **19**, 639 (2004).
11. M. Huber, Science **321**, 353 (2008).
12. B. M. J. Engelbrecht et al., Nature **447**, 80 (2007).
13. R. Condit et al., J. Trop. Ecol. **12**, 231 (1996).
14. R. Condit et al., Science **295**, 666 (2002).
15. K. B. Suttle, M. A. Thomsen, M. E. Power, Science **315**, 640 (2007).
16. In 15 arbitrarily chosen Rubiaceae, Colwell et al.'s elevation ranges are underestimated by an average of 800 m compared to Costa Rican collections (www.gbif.org).
17. C. Huntingford et al., Philos. Trans. R. Soc. B **363**, 1857 (2008).
18. F. Skov, J.-C. Svenning, Ecography **27**, 366 (2004).
19. J.-C. Svenning et al., Ecography **31**, 316 (2008).

GEOPHYSICS

Volcanic Symphony in the Lab

Luigi Burlini[1] and Giulio Di Toro[2]

Analysis of acoustic signals from lab samples links rapid pressure drops of pore fluids with low-frequency volcanic earthquakes.

Like philharmonic orchestras that perform symphonies with different musical instruments, active volcanoes produce a mix of seismic signals (earthquakes) that vary in their periodicity. Because each type of signal is associated to different physical processes, seismic monitoring can be a powerful tool for eruption forecasting, especially when combined with geochemical data (such as composition of escaping gases) and ground-deformation monitoring (1). The key issue is how to associate volcanic processes—which include fracture and dike propagation, magma feeding, and degassing—with each type of earthquake (2, 3). One approach is to recreate volcanic conditions with small laboratory samples, and then extrapolate the experimental signals (sonic to ultrasonic waves) to the scale of volcanic features. On page 249 of this issue, Benson et al. (4) measured acoustic emissions (AEs) in a basalt sample from Mount Etna

during loading and fracturing, and then on a rapid decompression of fluid. The AE signals recorded during the pore fluid decompression are similar to those detected during low-frequency earthquakes associated with volcanoes, which suggests that some natural quakes also originate from the rapid release of pressure in fluids (melts, gas, and supercritical fluids) flowing in fractures.

The seismic signals from volcanoes include high-frequency waves similar to those detected during tectonic earthquakes as well as low-frequency or long-period earthquakes and very-long-period earthquakes; tremors (continuous low-frequency ground vibration) and hybrid events that can mix these signals are also observed (2). Several theories have been proposed that connect volcanic with different seismic signals (3), but lab experiments potentially can allow observation of each physical mechanism separately—just as a clarinet passage is easier to recognize in a symphony performance if you have first heard the clarinet playing alone. Benson et al., using the tools typical of passive seismology (which looks at seismic signals and includes three-dimen-

sional earthquake location, waveform analysis, and computation of focal mechanisms), interpreted volcanic seismicity on the basis of experiments that reproduce variation of the physical conditions (such as pressure drop in a conduit) occurring in volcanic environments.

In rock deformation experiments, AEs are elastic waves produced by local strain events such as microfracturing, interaction of fluids with the crack walls, etc. (5, 6), although only rarely are emissions transmitted at audible frequencies (20 to 20,000 Hz). Following the pioneering work of Obert and Duvall (7), who used geophones to measure these signals, the use of arrays of piezoelectric transducers has enabled researchers to pinpoint the source of the AE and follow the evolution of sample damage (6, 8, 9).

To what extent can we link these lab studies and data from volcanoes? Experimental and natural waveforms can be similar in shape but can differ by orders of magnitude in frequency and amplitude (see the figure). Earthquakes are detected by seismometers and accelerometers that record ground motion and acceleration, whereas AEs are detected by

[1]Institute of Geology, ETH Zürich, CH-8092 Zürich, Switzerland. [2]Dipartimento di Geoscienze, Università di Padova, 35137 Padova, Italy. E-mail: luigi.burlini@erdw.ethz.ch; giulio.ditoro@unipd.it
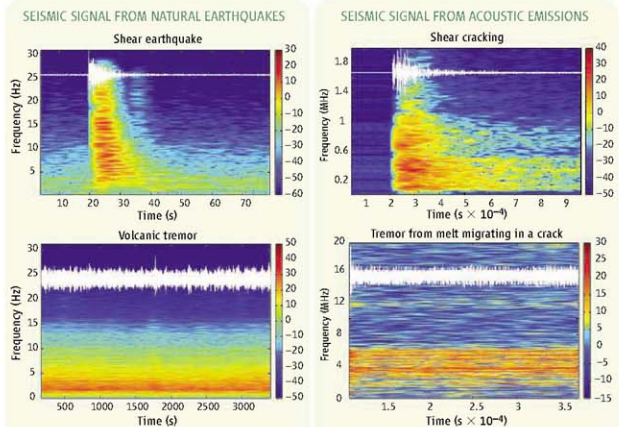
Scaling the volcanic symphony. Seismic signal (white line) and spectrogram from Mount Etna earthquakes (**left column**) and acoustic emissions (AEs) (**right column**). Note the similarities of natural shear earthquake with shear-cracking and of volcanic tremor with the signal from melt migration in a small sample. (**Top left**) Shear earthquake, 11 April 2002, 10:22. (**Bottom left**) Long-lasting tremor, 12 November 2002, 23:00. (**Top right**) AE of a shear crack produced during loading and fracturing a basalt from Mount Etna. (**Bottom right**) AE emitted from a basaltic melt intruding an ~0.1-mm-long fracture in an olivine aggregate. Color bars are arbitrary units from normalized amplitudes.

piezoelectric transducers, which record stress variation (and therefore deformation of the transducer and its acceleration). There are several reasons for believing that the lab studies are good models for earthquakes:

1) The ratio between the energy content of the AEs and their occurrence follows the Gutenberg-Richter earthquake statistic law (6, 10), which states that the cumulative number of events is inversely proportional to their energy.

2) There are strong similarities between the small quakes that follow a main earthquake (the so-called aftershock sequence) and the AE activity that follows sample fracturing (6).

3) AEs range from 10^5 to 10^7 Hz, whereas seismic waves have frequencies around 1 Hz. This difference can be related to the length of the fractures generated (at most 1 cm in the lab versus kilometers in the field). Benson et al. confirmed that the length of the microfractures observed after the experiments scaled with the dominant frequency of the AEs (11), similar to what is predicted for natural earthquakes (12).

4) The fluid viscosity of the fluid producing a tremorlike signal can be scaled from laboratory to nature, as Benson et al. have documented.

5) The spectrograms of natural and laboratory seismicity, once allowance is made for

the different frequencies, are similar. We infer that the evolution of the seismic signal with time is similar, which suggests that the underlying physical process is the same.

This final point is well illustrated in Benson et al.; the introduction of waveform analysis of the AEs allowed the synthesis of spectrograms that describe the frequency and amplitude content of each AE (11) that are remarkably similar to those of natural seismic waves (see the figure).

Experimentally reproducing sections of the volcanic symphony in the laboratory has several advantages but some limitations. Application of pressures and temperatures typical of volcanic areas at depths of 5 km (about 130 MPa and 300° to 600°C) requires small sample size (centimeter scale), and the AEs have small amplitudes because they originate from micrometer-scale displacements. Because the seismic signal decays rapidly with distance, the piezoelectric transducers must be placed as close as possible to the sample. Unfortunately, the transducer performance drops above 250°C, so working at higher temperatures requires that ceramic buffer rods are placed between the sample and the transducers to keep the latter cool.

Because Benson et al. were mainly looking at the effects of fluid–conduit wall inter-

actions, they could work at room temperature and place the transducers directly on the specimen. Their experimental configuration reproduces schematically a volcanic conduit, and their array of piezoelectric transducers simulates a volcanic seismic network. The waveforms Benson et al. recorded (often referred as microseismicity) were in fact similar to earthquakes registered during natural volcanic activity, despite working at room temperature. The microstructural analysis of the specimens after the experiments allowed them to identify a damage zone where the microseismicity was localized. From the geometry of the newly created fractures within the damage zone, they built a model explaining the nature of the microseismicity. Just as the turbulence of the air at the tip of a clarinet emits the typical sound, the tortuosity of the damage zone produces the turbulence of the fluid flow during the decompression. The fluids in volcanoes can be magma, supercritical fluids, and gases, but Benson et al. successfully simulated a volcano using only water.

Scale-invariant numerical modeling could be used to extrapolate these results to the dimensions of volcanoes and make the comparison more robust. This understanding should allow for better predictions of the intensity and timing of volcanic eruptions, so that early warning and alert can save lives.

References

1. H. Sigurdsson, Ed., *Encyclopedia of Volcanoes* (Academic Press, San Diego, CA, 2000).
2. R. McNutt, in *Encyclopedia of Volcanoes*, H. Sigurdsson, Ed. (Academic Press, San Diego, CA, 2000), pp. 1015–1033.
3. B. Chouet, *Pure Appl. Geophys.* **160**, 739 (2003).
4. P. M. Benson, S. Vinciguerra, P. G. Meredith, R. P. Young, *Science* **322**, 249 (2008).
5. M. S. Paterson, T. F. Wong, *Experimental Rock Deformation—The Brittle Field* (Springer, Heidelberg, ed. 2, 2005).
6. K. Mogi, *Experimental Rock Mechanics* (Taylor and Francis, London, 2007).
7. L. Obert, W. Duvall, *U.S. Bur. Min. Rep. Invest. 3803* (1945).
8. D. Lockner, *Int. J. Rock Mech.* **30**, 883 (1993).
9. J. S. Fortin, S. Standhis, G. Dresen, Y. Gueguen, *J. Geophys. Res.* **111**, B10203 (2006).
10. C. H. Scholz, *Bull. Seismol. Soc. Am.* **58**, 399 (1968).
11. L. Burlini, S. Vinciguerra, G. Di Toro, G. De Natale, J.-P. Burg, *Geology* **35**, 183 (2007).
12. K. Aki, P. Richard, *Quantitative Seismology* (University Science, Sausalito, CA, 2002).

INTRODUCTION

Lemons, Oranges, and Complexity

FACED WITH AN EPIDEMIC OF SCURVY, BRITISH SURGEON JAMES LIND DECIDED TO test some cures in 1747. In a legendary trial, he put 12 sickly sailors on a fixed diet and divided the group into twos. Each pair received a different supplement—cider, seawater, dilute acid, and so on. The result: After 6 days, the two who ate lemons and oranges improved so much that they were sent back on duty. The navy was slow to recognize the significance of Lind's work but eventually decreed that citrus (a source of vitamin C that prevents scurvy) must be in all sailors' rations.

By modern standards, Lind's test was weak. It had no control group. No placebo. No blinding of data. But it worked. By controlling the environment and varying treatments, it made possible an accurate reading of the human body's response. This idea, with many refinements, drives the modern clinical trial. But now the clinical trial itself is facing an epidemic—of rising costs and blurred objectives.

On page 210, David Malakoff reports on the view, as one U.S. researcher puts it, that the "system isn't working." Malakoff describes factors that are driving inflation. Among them are complex rules for reporting, the proliferation of required tests, the difficulty of recruiting and keeping volunteers, and waste in the design and management of trials. But there is an encouraging note, too: Public and private organizations are aware of these problems and are pushing for reform.

One counterpoint to problems in the West is a boom in the East. Dennis Normile describes the expansion of clinical studies in Asia on page 214. A desire to cut costs is only part of this trend; China and India see clinical trials as a means of growing their economies and developing skills.

Jocelyn Kaiser reports on page 217 about ongoing efforts to make Western trial data more accessible. A mandate requiring disclosure of results from all trials of drugs approved by the U.S. Food and Drug Administration takes effect this fall, for example. On page 219, Constance Holden gives an update on efforts to ensure that U.S. government–sponsored trials are balanced by gender. And Jennifer Couzin dissects an important series of interventions designed to prevent heart disease by controlling cholesterol (p. 220). The results raise fundamental questions about how cholesterol functions in the body—and about the usefulness of certain biomarkers as surrogates for health outcomes.

A video report by Robert Frederick on progress in testing drugs for children, "Pediatric Medicines: Prescribing Off-Label," appears online at www.sciencemag.org/clinicaltrials.

The subtlety and cost of these modern clinical trials would amaze James Lind.
                                                                    –ELIOT MARSHALL

Clinical Trials

CONTENTS

News

See also related online material at www.sciencemag.org/clinicaltrials
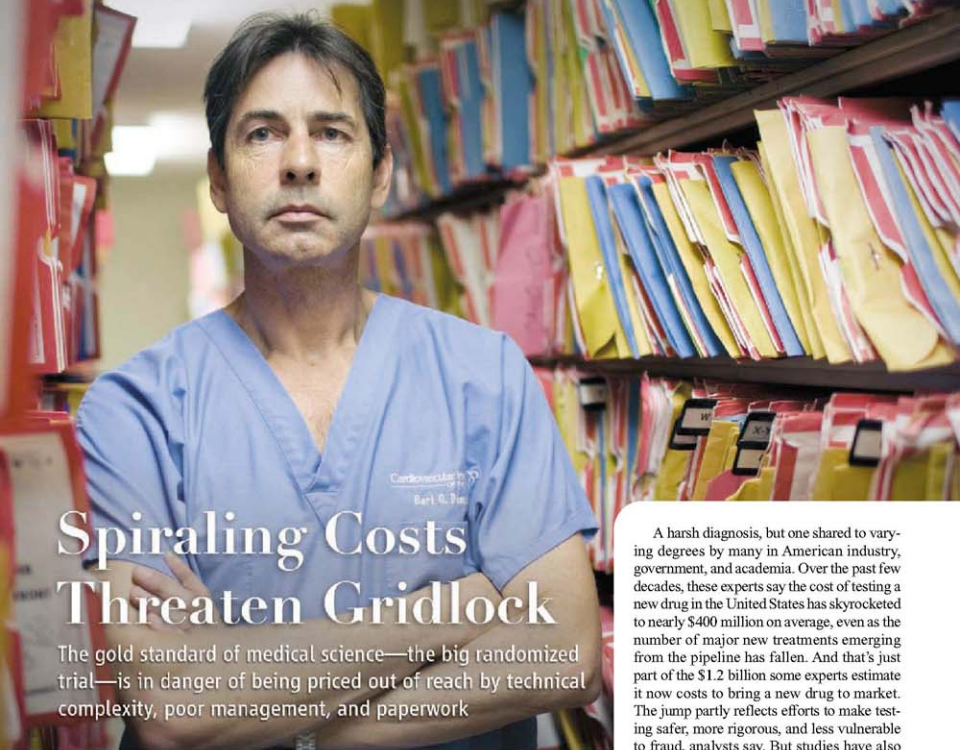
Science

Spiraling Costs Threaten Gridlock

The gold standard of medical science—the big randomized trial—is in danger of being priced out of reach by technical complexity, poor management, and paperwork

CARDIOLOGIST BART DENYS SAYS THAT IF you want to understand why testing new medical treatments can be so discouraging and so expensive these days, just leaf through the proposed clinical trials that pour into his office. Pharmaceutical and medical-device companies seek him out in Thibodaux, Louisiana, because he's an experienced clinical researcher who works for a private institute with access to more than 30,000 patients.

The companies need help from those patients to run trials on everything from experimental heart drugs to better implants—and to win government approval to sell their products. They are willing to pay physicians and subjects to participate. But the process "can be very frustrating," says Denys, who has done clinical studies for more than 20 years and is now director of clinical research for the Cardiovascular Institute of the South, a large private practice.

Many trials, for instance, require Denys to wrangle piles of paperwork and guide patients through a maze of lengthy consent forms. Some come with enrollment criteria that are so stringent it's nearly impossible for him to find subjects who qualify. Others require volunteers to undergo dozens of time-consuming and sometimes painful medical procedures. Then there are the occasional studies that Denys dismisses as "hidden marketing campaigns," trials that appear designed more to promote the use of a product than to answer important questions.

The take-home message for U.S. biomedical researchers? "The system isn't working," Denys says. "We're wasting too much time and money on trials that are poorly designed and difficult to execute. They take too long. They produce trivial information and not enough important treatments. They don't ask relevant scientific questions. Clinical trials are broken, just broken."

A harsh diagnosis, but one shared to varying degrees by many in American industry, government, and academia. Over the past few decades, these experts say the cost of testing a new drug in the United States has skyrocketed to nearly $400 million on average, even as the number of major new treatments emerging from the pipeline has fallen. And that's just part of the $1.2 billion some experts estimate it now costs to bring a new drug to market. The jump partly reflects efforts to make testing safer, more rigorous, and less vulnerable to fraud, analysts say. But studies have also grown more complex because researchers are testing more sophisticated therapies. Also to blame are poor corporate planning and regulatory chaos.

Analysts fear the trend is imperiling drug development by alienating research foot soldiers such as Denys and driving clinical science out of the United States to nations with laxer rules. "Parts of the system are starting to collapse, and that doesn't inspire confidence," says Kevin Schulman, who studies the economics of clinical trials at Duke University in Durham, North Carolina. "We could wake up one day and discover that investors don't want to put money into drug R&D in the U.S."

Still, clinical research insiders say it's not too late to fix the problems. Academic centers and government agencies, including the U.S. Food and Drug Administration (FDA), have launched initiatives aimed at improving the efficiency of the thousands of trials begun each year. Sponsors are already trying simpler

> **Clinical blizzard.** Cardiologist Bart Denys says it's getting harder than ever to conduct clinical trials in the United States, partly due to growing record-keeping requirements.

designs and automated record-keeping. Some companies even want to scrap the traditional three-step approach to trials.

"No question, these are challenging times, but this is an amazingly resourceful and innovative enterprise," says Kenneth Getz, a prominent clinical research analyst at the Tufts Center for the Study of Drug Development in Boston. "Time and time again, the pharmaceutical industry has shown it's not going to let itself disappear."

Rising expectations

Many of the 60,000-plus clinical trials registered by the National Institutes of Health on its ClinicalTrials.gov Web site (see p. 217) follow a familiar but not formally required three-step process. Phase I trials typically test for safety in a few dozen volunteers. Phase II studies expand the testing to several hundred subjects, usually with the target condition, to gather more data on safety and preliminary evidence on whether the treatment actually works. If phase II is successful, phase III trials enroll thousands or even tens of thousands of patients at dozens of sites around the nation—and increasingly around the world—to measure effectiveness, with researchers watching closely for side effects.

In the past, industry conducted most of this work at university medical centers, which had plenty of patients and physicians interested in

research. In the past decade, however, many factors, including cost concerns, have prompted trial sponsors to work more with private doctors such as Denys. "Just 15, 20 years ago, most trials were pretty straightforward," he says. "I could find 500 patients to enroll in a phase III drug trial without much problem." The paperwork was pretty brief, and the data on a single subject might fill just a single page. "Not anymore," Denys says in a wistful tone. "It's gotten a lot more complicated."

Several recent studies suggest he's right. A team led by Getz, for instance, has been combing through a massive trove of trial data compiled by Medidata Solutions Worldwide, a New York City–based company that consults on clinical research. After looking at protocols for more than 10,000 trials approved between 1999 and 2005, the researchers concluded that studies now require more time and effort than ever from doctors such as Denys.

One time sink, Getz concludes in a study published in the May issue of *Regulatory*

Trials by the Numbers

Enrolling more subjects (average)
1985: 1700
2005: 4200+

Harder to recruit subjects (average)
1999: 10 enrollment criteria
2002: 26 enrollment criteria

Fewer subjects finishing
1999–2002: 69%
2003–2006: 48%

Few drugs make it
Compounds entering phase I trials ultimately approved for sale:
5% to 10%

More procedures (median)
1999: 96 per trial
2005: 158 per trial

Taking longer to complete
1999–2002: 460 days
2003–2006: 780 days

Affairs Journal Pharma, is the growing number of enrollment criteria. A patient might have to be within a certain body weight or age range, for instance, or not taking certain other medications. Between 1999 and 2005, the median number of criteria in a typical trial jumped dramatically, from 31 to 49. In part, this reflects an industrywide shift to more nuanced drugs focused on more complex diseases, Getz believes, and greater industry attention to specific demographic groups. "That's not necessarily a bad thing," he says. "But the end result is that it is getting a lot harder to find patients you can 'rule in' to a trial."
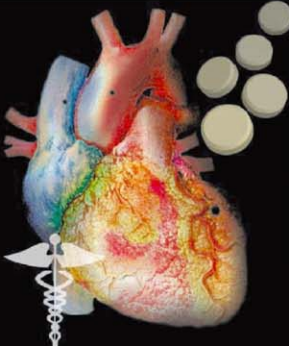
At the same time, the study found that patients who do enroll in trials can count on being poked and prodded more than ever. In 1999, the typical study involved 96 total procedures, such as blood tests or electrocardiograms. By 2005, the list had grown more than 50%, to 158.

That trend is being driven by both scientific and regulatory factors, Getz and other researchers say. In part, trial designers are taking advantage of a host of new

> **Hard to explain.** Trials are getting so complicated that it's hard for physicians such as Denys to lay out the potential risks and benefits for patients.

Clinical Trials and Tribulations

Say it's late 2001, and you want to compare several drug-based treatments for heart disease in a double-blind, phase III trial enrolling 14,500 patients. Get ready to spend at least $102 million to $207 million, according to a 2005 study led by medical professor Eric Eisenstein of Duke University in Durham, North Carolina. Six veteran pharmaceutical executives developed those estimates, and Eisenstein says it would probably cost more today. About two-thirds of the cost is in monitoring visits to sites and management.

The Trial

Subjects: 14,500
Trial sites: 800
Length: 46 to 71 months
Pages per final patient report: 22 to 134
Monitoring visits per site: 6 to 18

Big-Ticket items:

Site and subject payments: $34 to $102 million (48%)
Site management: $31 to $53 million (30%)
Data management: $11 to $21 million (11%)
Safety reviews: $0 to $9 million (3%)

Total: $102 to $207 million

$ Economy option

Limit trial to 60-page reports and 10 monitoring visits.

Total: $140 million

$ Supersaver

Limit trial to 10-page reports and four monitoring visits

Total: $85 million

technologies, including high-tech scanners and sensitive genetic tests, that can give them a better idea of how an experimental treatment is working. But analysts say they are also ordering needless procedures in order to be ready for questions from regulators. "Nobody ever looks at a lot of this data. It's totally wasted," says Robert Califf, vice chancellor for clinical research at Duke. "They just want to have it because they perceive it might come in handy if something comes up."

Ironically, however, Tufts researchers say the extra work isn't producing higher payments to doctors and patients. They say these fees, which can amount to several thousand dollars per subject for the doctor and hundreds of dollars for the patient, have been flat or declining over the past decade.

"You're beginning to see this huge mismatch: The clinical researcher and patients are being asked to do more for less," says Medidata's Ed Seguine, who worked with Tufts on the study.

Battalions of checkers

The slumping fees, however, don't mean trials are getting cheaper. Other studies are finding that growing complexity is having a profound impact on bottom-line costs, sometimes in hidden ways.

Difficulties in enrolling patients, for instance, mean that trials often take far longer than planned. The Tufts analysis found that trial length rose by 70% between 1999 and 2006, to an average of 780 days. "Delays are costly because time is literally money in this industry," says Kenneth

Kaitin, head of the Tufts center. Among the problems: Delays reduce the value of intellectual property, which has a limited life; give competitors an edge in the race to the market; and increase borrowing costs. Overall, Tufts researchers estimate that "time costs" now account for roughly half of the cost of getting a new drug approved in the United States.

Complexity also drives up the cost of scrubbing data. Sponsors employ a small army of monitors who visit trial sites and examine and correct faulty databases—sometimes at costs of up to $350 per data point, according to industry veterans. "Just sticking another procedure on the form might not seem like much, but it can have big downstream ripple effects," says Califf.

One strategy trial sponsors are using to reduce costs is to enlist the help of experienced investigators such as Denys, who have access to large pools of patients. Such veterans are in short supply, says Christine Pierce of RxTrials, an Ellicott City, Maryland–based clinical research company that has run hundreds of trials. "Most physicians are simply not involved" in research, she says, noting surveys that suggest less than 15% of U.S. doctors participate in clinical studies, and those that do typically drop out after just a few years. "They find out it's burdensome, difficult, and not particularly lucrative," she says.

As a result, many trials rely on neophyte clinicians who, despite their best intentions, have difficulty recruiting even a few patients. Indeed, some companies now assume that up to a quarter of the "sites" in a study—anything from a fully equipped academic medical center to a solo doctor's office or temporary storefront—will never enroll a single patient. Those false starts "get very expensive very quickly," says Pierce, noting that it can cost $20,000 to qualify a site for a trial.

Slimming down

Those numbers are one reason drug- and devicemakers are increasingly moving trials out of the United States to China, India, and Eastern Europe (see p. 214). "They believe they can enroll more patients faster and that the subjects will stay in the trials," says Califf. But he says companies won't completely abandon U.S. trials any time soon, if only because FDA is unlikely to approve treatments for use here unless they've been tested on at least some Americans. So "there's a lot of interest in figuring out some new ways of doing this," he says.

He hopes that at least some of those solutions will come from the new Clinical Trials Transformation Initiative, begun last year by Duke and FDA. Among its targets: Finding ways to standardize and automate paperwork, identify and accredit experienced clinical researchers and sites, reduce monitoring overhead, and clarify what kinds of data are essential for regulators. "The clinical research enterprise needs to evolve," Janet Woodcock, FDA's deputy commissioner, said in unveiling the initiative last November. "It needs to be much more streamlined and efficient."

Those goals also underpin another, longer running FDA effort, called the Critical Path Initiative. Since 2004, the agency has been trying to tackle a raft of trial-related issues, including how best to deal with incomplete trial data and how to promote innovative, "adaptive" trial designs. Such studies allow investigators to peek at trial data before the study is formally over, potentially allowing them to shorten trial times or make midcourse corrections.

Many companies aren't waiting, however. "I'm seeing all kinds of experiments," says Getz of Tufts. At pharma giant Wyeth, for instance, executives are moving to replace the traditional three-phase trial process with an approach called "Learn and Confirm." Based on ideas promoted by the late biomedical researcher Lewis Sheiner, it splits the testing process into two phases overseen by different teams of researchers. The learn team leads the equivalent of phase I and II studies, while a confirm team tries to finish the process. Wyeth executives hope the competitive reviews will lead to leaner, meaner trial designs and fewer wasteful late-stage trials driven by researchers unwilling to kill off pet projects.

Other companies are selling off in-house research units and turning to outside contract research organizations in a bid to cut management costs, which often account for more than half a study's price tag. Studies suggest it's an area ripe for cost-cutting (see graphic, p. 212). They're also hiring what one industry insider calls "site whisperers," consultants who use massive databases and extensive personal knowledge to identify physicians and sites likely to produce patients and high-quality data.

Companies are also hiring consultants to analyze trial complexity. Seguine, for instance, tells how he showed a company that one of its trials was about three times more complex than the standard. "Complexity isn't necessarily good or bad," he says. "But you need to consider it."

In Louisiana, that's just the kind of advice Bart Denys wishes more trial sponsors were getting. "Some of the studies I see are so complicated I wouldn't be able to enroll a single patient," he says, much less concisely explain the risks or study design. Others, he adds, push the boundaries of what's ethically acceptable. He refuses, for instance, to participate in what he believes are "seeding studies" designed to acquaint doctors with a product (see sidebar, below). And he also dislikes studies that may provide little benefit, and possibly great misery, for volunteers. "I saw one recently that would have required me to inject a placebo into a patient more than 100 times in just a few weeks. I mean, who is designing these things? It's pretty obvious they don't see patients."

Still, Denys says he's "hanging in with clinical research because it is intrinsic to how I practice medicine—I want to answer questions." The question now is whether that kind of curiosity will, in the long run, be enough to sustain clinical research in the United States.　**–DAVID MALAKOFF**

David Malakoff is a writer in Alexandria, Virginia.

Allegations of Waste: The 'Seeding' Study

In March 1999, 600 doctors began enrolling more than 5500 arthritis patients in a short clinical trial aimed at comparing the gastrointestinal safety of the new Merck painkiller Vioxx with an existing treatment. But internal company documents suggest the trial had little to do with science and much to do with marketing, according to a recent analysis. Merck denies the charge, but the study has focused new attention on so-called seeding trials aimed at promoting new treatments.

The documents were obtained during the discovery process by attorneys suing Merck over Vioxx's safety. They suggest that Merck's marketing division "designed and executed" the 1999 ADVANTAGE trial, which at least one top company scientist criticized as "intellectually redundant" and "wasteful," physician Harlan Krumholz of Yale School of Medicine and three colleagues reported in the 19 August issue of the *Annals of Internal Medicine*. The four authors, who are paid consultants to attorneys suing Merck, also claim that Merck employees routinely referred to ADVANTAGE as a seeding trial, "but let's not call it that in our internal documents," one wrote in an e-mail, according to the *Annals* article. The company's goal, the authors conclude, was to lure leading physicians into the habit of prescribing Vioxx in the months leading up to the U.S. Food

Sowing seeds. Critics say Merck developed one clinical trial primarily to promote, not test, its painkiller Vioxx.

and Drug Administration's approval of the drug—although the purpose of the trial was not made clear to patients or doctors. Although seeding trials may be a longtime open secret in the industry, the authors write that the Merck documents "provide the first strong documentary evidence" of the practice.

Merck has vigorously defended ADVANTAGE. "The primary intent of the study was to answer scientific questions of importance to primary care physicians," Jonathan M. Edelman, executive director of Merck's Global Center for Scientific Affairs, wrote in an open letter. Merck's business interests "were clearly understood" by independent reviewers who approved the trial, he added. And the *Annals* itself, he notes, published ADVANTAGE's results in 2003 because editors thought the findings were useful.

Annals Editor Harold Sox, however, now says the journal would not have published the results had it known of the trial's apparent purpose. "Deception is the key to a successful seeding trial," Sox and *Journal of the American Medical Association* Deputy Editor Drummond Rennie conclude in a recent editorial, adding that it's up to reviewers, doctors, and even patients to ask questions that might reveal when promotion is masquerading as research.　**–D.M**

The Promise and Pitfalls of Clinical Trials Overseas

Big pharma has big incentives, including cost savings and more powerful studies, to launch trials in developing countries. But can companies avoid the ethical potholes?

THE RESULTS OF THE CLINICAL TRIAL WERE puzzling. Some lung cancer patients who received the experimental drug gefitinib several years ago showed almost no benefit; in other patients, tumors shrank so much that one researcher called it a "Lazarus type of response." After intense study, an answer to the riddle emerged: Tumors that respond to gefitinib have a mutation in a key protein affecting cell growth—a mutation common in Asians but rare in other races.

As a result, gefitinib, which is marketed as Iressa by AstraZeneca, is available only under special circumstances in North America, while in Asia it has become an established therapy for non–small cell lung cancer that fails to respond to other treatments. If AstraZeneca had done the initial trials in a global setting, "they might have made the Asian connection sooner and saved a lot of money and time," says Benny Zee, a biostatistician and director of the Comprehensive Cancer Trials Unit at the Chinese University of Hong Kong.

Differing ethnic responses to drugs is one of a host of reasons pharmaceutical companies are globalizing clinical trials—and rushing to developing countries. There are cost savings. There are new markets. And trials help a multinational pharmaceutical company establish a presence in a country and learn local needs.

Large pools of recruitable subjects and enormous market potential have drawn drug companies to India and China. A July study by the Associated Chambers of Commerce and Industry of India (ASSOCHAM) foresees India's clinical trials business growing from less than $150 million currently to $546 million by 2010. Among developing countries, India and China are possibly the best positioned to leverage multinational clinical trials to develop their own pharmaceutical industries, thanks to their rapid economic development and growing scientific capabilities.

But bioethicists worry that in the stampede to Asia, patients' rights sometimes get trampled. "All those reasons [for doing clinical trials in India] don't really add up to good ethical oversight," says Prathap Tharyan, a psychiatry professor at Christian Medical College in Vellore, India. Tharyan says India should view the boom in trials as an opportunity to raise ethical standards.

In the latest flap sparking calls for closer scrutiny of trials, the All India Institute of Med-

ical Sciences in New Delhi acknowledged in response to a query from a nongovernmental organization in August that 49 infants died while enrolled in clinical trials at the institute over the past 2.5 years. "The deaths were among patients who were extremely sick; there is no case of a death because of an intervention," says institute spokesperson Y. K. Gupta. All trials, he says, were vetted by an ethics committee and conducted in accordance with good clinical practice. In early September, the institute reported the results of an in-house investigation to the health ministry, but the matter may not end there.

Cheaper, faster

Reliable statistics on who is doing which trials, and where, are hard to come by. Few countries require the registration of clinical trials in public databases, and individual trials are becoming more and more global. For drugs for the U.S. market, the lead researcher at each site must file a statement with the U.S. Food and Drug Administration as part of an Investigational New Drug application. According to an analysis of filings by CenterWatch, a Boston-based company that gathers data on clinical trials, the number of investigators in developing countries working on drugs for the U.S. market has risen dramatically: in India, from 46 in 2001 to 493 in 2007, and in China, from 16 to 97 over the same period. (The CenterWatch analysis captures a fraction of ongoing trials.) Also telling, the ASSOCHAM study found that India had been the site of a single clinical trial outsourced by U.S.-based multinationals from 1996 to 2000—but 192 trials between 2001 and 2005.

One factor behind this trend is economics. The Confederation of Indian Industry boasts that trials in India can be 50% to 60% cheaper than in the United States. But cost advantages are narrowing, says Kenneth Kaitin, director of the Center for the Study of Drug Development at Tufts University in Boston. "These

Queuing up. Clinical trials are coming to India—and elsewhere in Asia and Eastern Europe—drawn by low costs and diverse populations.

countries realize there is no reason for them to reduce costs to the degree they have in the past," he says. Still, costs are low by Western standards (see p. 210).

Taking trials overseas can pay off in another way. The cost of running a trial "is a factor to some degree, but not to the degree that people think," says Jorge Puente, vice-president for medical and regulatory affairs in Japan and Asia for Pfizer Inc. in New York City. He explains that once a drug is patented—typically before trials begin—the clock starts ticking on the period of exclusivity. Trial sites in North America and Europe already have hundreds of ongoing studies, so competition for patients delays recruitment and a trial's completion. Target patient numbers can be gathered more quickly if trials include sites in developing countries. "If you speed up development by 1 year, you get an extra year of [patent] exclusivity; that's the most important driver," Puente says.

It's in the genes

Liver cancer and gastric cancer kill more people in China every year—600,000—than the number who die from all forms of cancer in the United States, Puente says. Developing drugs against these and other diseases that afflict China disproportionately is the main reason pharmaceutical companies have opened R&D centers in China (*Science*, 27 July 2007, p. 436). The same goes for India, where the focus is on developing drugs for malaria and other infectious diseases seldom seen in North America and Europe. "We have to be part of those communities to understand the health priorities and be part of the solutions," says Puente. He says that more than 60% of Pfizer's revenue now comes from outside the United States, and projections indicate an additional billion people in Asia will be potential consumers of innovative medication.

The buildup of research infrastructure in China and India is leading to "a more strategic approach" to trials, Kaitin says. In the past, overseas trials were primarily part of large phase III studies. "Now companies

may do a phase II study to reach proof of concept more quickly so they can determine whether to move forward, and more Western companies are looking to conduct preclinical and research discovery phases of drug development in these countries," he says.

As the gefitinib experience shows, ethnic groups often respond to drugs in markedly different ways. Genomic data will allow the medical community to better understand and exploit differences, Zee says. "Before talking about personalized treatment, we can talk about the genetic makeup of different ethnic groups," he adds.

Indian scientists are touting the country's diversity as an advantage. The Indian Genome Variation Consortium is studying genetic differences among India's ethnic groups and comparing them to variations in other populations. Preliminary results covering 55 populations and several hundred genetic markers were reported in the April 2008 issue of the Indian

Academy of Sciences' *Journal of Genetics*. Lead investigator Samir Brahmachari, a genome analyst and director general of the Council of Scientific and Industrial Research in New Delhi, says India offers a one-stop destination for clinical trials, because so many human variations exist in the 1-billion-plus, drug-naïve population.

Courting trials

A final factor that's fueling the boom in clinical trials in developing countries is their active solicitation—a sharp turnabout from a mere decade ago. "In the early to mid-'90s, there really was no infrastructure for clinical research in Asia," says Puente, who worked for Pfizer in China and Japan at that time. He says few researchers were interested. These days, China and India are promoting themselves as the places to be for clinical trials. The ASSOCHAM report boasts that India churns out 17,000 new doctors each year, all of whom speak English. Kaitin says conducting trials brings money and staff training to hospitals, medical schools, and local research organizations.

Both countries also have domestic companies eager to move from making copycat generics to developing their own drugs. At the East Asian Pharmaceutical Regulatory Symposium in Tokyo last April, Zhang Wei, director-general of the Department of Drug Registration at China's State Food and Drug Administration (SFDA), reported that from 2001 to 2005, China's domestic drug companies had received approval for 45 new drug products and had 41 under review and 109 in various stages of clinical trials. (The U.S. Food and Drug Administration approved more than 50 new drugs in 2007 alone.) China's government has launched several schemes in recent years to promote drug development, including a National Key New Drug Creation Program approved by the cabinet last December that promises to make up to $1.5 billion available to academics and biotech start-ups over the next 15 years. China and India are working to streamline approvals processes, crack down on corruption, and improve intellectual-property protection. "These countries are really doing whatever they can to become major players

Trials in Developing Regions
(% of total filed with FDA)

Legend:
- Middle East/Africa
- South America
- Central/Eastern Europe
- Asia/Pacific

2003: 8.9% (Asia/Pacific 2.1%, Central/Eastern Europe 3.8%, South America 2.0%, Middle East/Africa 1.0%)
2004: 11.4% (2.7%, 5.0%, 2.2%, 1.5%)
2005: 12.8% (2.9%, 6.1%, 2.5%, 1.3%)
2006: 14.7% (4.1%, 6.2%, 2.9%, 1.5%)
2007: 17.54% (4.2%, 7.9%, 3.6%, 1.9%)

Trials in Key Countries
(number filed with FDA)

Legend:
- Russia
- Brazil
- India
- China

Russia: 189 (2002: 208), 307, 496, 628, 713, 674
Brazil: 96/195, 235, 271, 235, 271, 281
India: 46/43, 105, 174, 225, —, 493/390
China: 16/14, 26, 76, 72, 87, 97

in this expanding global market," says Kaitin.

Still, China and India are held back by an underdeveloped infrastructure. Zee, who studied and worked for more than 15 years in the United States and Canada and whose research focuses on cancer radiation therapy trials, says having lead scientists trained in North America and Europe is not enough. "You need technicians; you need good laboratory and other [supporting services] to do a trial properly," he says. China's SFDA allows new drug and device trials at only 200 or so approved centers—a "terribly small" number given China's size, Puente says. "The limiting factor for us," he says, is the lack of SFDA-certified centers.

The biggest infrastructure gaps in both countries are in trial know-how and ethical oversight. Wu Taixing, an epidemiologist at Sichuan University's West China Hospital in Chengdu, says only a few hundred of the top hospitals have ethics committees. Wu and colleagues found that just 207 of 2235 "randomized" trials reported in Chinese journals were randomized properly. "Most authors of these reports lack an adequate understanding of rigorous clinical trial design," Wu's group concludes. In India, too, "a large majority of potential investigators lack knowledge of regulations, ethics and good clinical practice, and skills for clinical trial management," says Arun Bhatt, president of ClinInvent Research India, a contract research organization in Mumbai.

Many trial sponsors use ethically dubious recruiting practices, such as offering payments that dwarf participants' normal earnings and providing medications they could not otherwise afford, alleges C. M. Gulhati, editor of the *Monthly Index of Medical Specialties*, a reference work on drugs published in New Delhi. Widespread illiteracy makes it easy to sidestep informed-consent procedures, Gulhati adds: "Investigators frequently enroll patients in trials as if their participation were a necessary next step in their care."

Critics have pounced on such lapses. "There are good reasons to believe that the rights of test subjects in developing countries are less secure than those of their counterparts in the West," according to a January report, *A Bitter Pill*, from the Wemos Foundation, an organization in Amsterdam that advocates for developing-world health care.

Among several examples the Wemos report mentions is a case in which a drug-eluting stent developed by Occam International in Eindhoven, Netherlands, was implanted in about 70 Indian patients as part of a 2005 trial. According to a 2006 report by the Netherlands Health Care Inspectorate, the procedure was done without proper informed consent and without ethics committee approval. The Inspectorate slammed Occam for its "amateurish" reliance on a local partner without verifying that the trial would be conducted in accordance with European ethical standards. In a statement responding to questions from *Science*, Biosensors International in Singapore,

Homegrown. Some products such as antiviral vaccines may be designed, tested, and marketed in the developing world.

Occam's parent company, said Occam has denied the accuracy of the allegations and noted that investigations in both India and the Netherlands produced "no finding of legal wrongdoing."

Wemos called for authorities in developed countries to adopt stricter controls to prevent drugs tested unethically from reaching the market. It also urged developing countries to beef up health care systems and ethical review capabilities and asked drug companies to be more transparent in reporting on clinical trials.

Many researchers emphasize that the issue transcends big pharma. "Pharmaceutical [companies] aren't the only ones that do clinical trials; we're looking at academic researchers, device companies, and public

health researchers as well," says Davina Ghersi, coordinator of the World Health Organization's International Clinical Trials Registry Platform (ICTRP), established in August 2005 as a one-stop portal for trial registries around the world.

To subject trials to greater scrutiny, groups in China and India launched registries last year that ask for information such as who is paying for a trial, the health issue being studied, the intervention, and the outcomes. These are the first registries in developing countries; both are cooperating with ICTRP. "Trial registration suggests an opportunity to help facilitate the ethical design, scientific design, and good conduct and reporting of trials," Tharyan says.

In addition to the minimum data required by ICTRP, Clinical Trials Registry–India asks registrants to provide details such as the method of randomization and blinding. The Indian registry also asks for information on the ethics committee that approved the trial. It hopes to evaluate committee qualifications and performance and eventually to accredit them. The Chinese Evidence-Based Medicine Center, one of the sponsors of the Chinese Clinical Trial Register, plans to help institutions set up and train institutional review boards.

Neither China nor India requires trials to be registered. To encourage registration, journal editors in both countries are following the lead of the International Committee of Medical Journal Editors, which announced 4 years ago that member journals will consider a trial for publication only if it is in a publicly accessible registry before enrollment of the first patient. Last February, editors from 12 of India's top medical journals announced a similar requirement to take effect in January 2010. In China, a consortium of 56 journals has announced that as of January 2009, reports of registered trials will get priority for publication over unregistered trials. Eventually, they will publish results from only registered trials.

Wu and Tharyan acknowledge that registries and journal policies won't resolve all ethics issues, but they are a start toward helping ethical oversight keep up with the developing world's growing participation in clinical trials.

–DENNIS NORMILE

With reporting by Pallava Bagla in New Delhi and Chen Xi in Beijing.

Making Clinical Data Widely Available

Granting public access to drug trial results and sharing patient data among researchers will make products safer and advance science, proponents say

LAST SUMMER, STATISTICIAN ANDREW Vickers e-mailed the lead author of a paper on a large clinical trial to ask for the underlying raw data set. Vickers wanted to explore whether he could predict which patients were at risk for stroke, a side effect of the drug being tested. He got a short reply, one he has heard many times before: The data are not available. "It's just the knee-jerk refusal to share data," says Vickers, at Memorial Sloan-Kettering Cancer Center in New York City. Vickers wasn't surprised by the response or the lack of explanation. But he's become so frustrated by such experiences that he penned an essay in the *New York Times* in January that called for making data from cancer trials freely available. Clinical researchers should get over their reluctance to share, according to Vickers; he and some others say they should emulate basic researchers who routinely deposit gene sequences in GenBank. "We need an attitude shift," Vickers says.

Change is happening, in fact, on two fronts. Starting last month, a law enacted in 2007 will require sponsors of all clinical trials of drugs and devices that were subsequently approved in the United States to post summary results in a federal database. The aim is to ensure that all findings see the light of day, including negative results that often get buried. At the same time, the National Institutes of Health (NIH), which ramped up its data-sharing efforts with genome data more than a decade ago, has been advancing these policies into clinical research. Some researchers are wary. They warn that clinical and epidemiological data sets can be difficult to analyze without help from the investigators who designed the study and understand its nuances. Totally free access to data, they argue, could lower the quality of analysis. Sharing raw data remains a touchy issue, and even the mandate to release summary results is stirring controversy.

This new policy, a provision in the 2007 Food and Drug Administration (FDA) law, expands a 1997 requirement that companies register trials of drugs to treat serious diseases at a public Web site (ClinicalTrials.gov). Adding basic data on outcomes and statistical analyses for all approved drugs and devices will allow outside researchers to analyze both positive and negative results, says Peter Lurie of Public Citizen, a consumer safety watchdog group. For example, a court-mandated database of summary data from GlaxoSmithKline trials led to a meta-analysis last year questioning the safety of the diabetes drug Avandia. "It's a corrective against misleading presentations of data," says Lurie.

It may be bare-bones data, but the new mandate will "represent a sea change" for some academic researchers, because the data must be posted within 12 months of a trial's completion or within 30 days after FDA approval—a shorter window than many researchers now take to publish, says Deborah Zarin of NIH's National Library of Medicine, who oversees ClinicalTrials.gov and discussed the changes in *Science* earlier this year (*Science*, 7 March, p. 1340). "It will incentivize people to get a publication out" before someone else publishes on the data, she says. As for drug companies, they "have always been supportive" of disclosing results for approved drugs and so far are satisfied with the implementation, says Alan Goldhammer, deputy vice president for scientific and regulatory affairs at the Pharmaceutical Research and Manufacturers of America.

More contentious is a proposal in the new law to extend the policy to unapproved drugs in 2010 and add narrative and lay summaries. "It's a lot of work [to prepare the data summaries], and we don't see that this information is particularly useful" to the public for drugs that can't be prescribed, Goldhammer says. Others worry that the narratives could preempt publication in a journal, and NIH's posting of them could be viewed as giving a drug "a stamp of approval," says Zarin.

Summary tables can only get you so far, in any case. Raw data on individual patients (stripped of names, addresses, and other identifying data) would be better, say Vickers and other

Free the data. Statistician Andrew Vickers thinks clinical researchers should be as willing to share their raw data as genome sequencers are.

academic researchers. Only the full data set allows others to verify the conclusions of a publication, notes Christine Laine, senior deputy editor of the *Annals of Internal Medicine*. This can save lives. Vickers points to a study led by two NIH intramural researchers 6 years ago on using proteomic signatures to screen for ovarian cancer. The test was poised to enter clinical practice until outside groups reanalyzed the data, which were posted on an NIH Web site, and questioned its validity. The full data set may also be needed for detailed meta-analyses. Vickers has also asked for raw data to explore new questions, design new studies, and test novel statistical methods.

Admittedly, clinical research data are more difficult to share than gene sequences, because investigators must remove identifiers and obtain patients' consent or a waiver, as required by U.S. privacy rules. Moreover, clinical investigators are understandably reluctant to hand over data sets in which they have invested years and that they hope will generate many papers. "They don't want to be scooped," says Laine.

Despite the obstacles, some clinicians have begun routinely sharing brain-imaging data (*Science*, 4 April 2003, p. 43). Many journals also require that gene expression data, including clinical results, be deposited in a public database.

Perhaps the broadest mandate for sharing patient data is one NIH announced in 2003 for all grants more than $500,000, which covers many human studies. The National Cancer Institute clinical trials groups were among the first to craft formal data-sharing plans. Other NIH institutes have set up repositories that disburse so-called limited-access data sets from clinical and population studies, typically within 2 years after the trial ends. You can't just download the data; requestors may have to submit a proposal, document approval from their local ethics board, and sign an agreement limiting how they use the data. Often there is no fee, although one cancer group says it charges a substantial $5000 on average per data set.

Although NIH is not tracking overall requests, one institute that was ahead of the curve, the National Heart, Lung, and Blood Institute (NHLBI), says its data-sharing program has been a success. The institute received nearly 100 requests for the 60-some available data sets last year, mostly for epidemiology studies but some for clinical data. Since 2000, at least 135 papers have resulted from such requests. Although that is modest compared to the hundreds of papers that the study leaders themselves have likely produced, NHLBI Director Elizabeth Nabel sees "a change in the culture. I think investigators are now much more open to data sharing and collaborative work."

Going a step further, in January, NIH began requiring all grantees conducting so-called genome-wide association studies, in

which patients' DNA is scanned for markers linked to disease, to share data or explain why they can't. The quality-checked data should be deposited in NIH's dbGaP database straight from the lab; the study investigators get a year's exclusive use for publications. Requests for data sets available since last year are pouring in, says Nabel—more than 50 investigators have asked for data from the famed Framingham Heart Study alone. NIH is now considering whether to extend its 2003 data-sharing policy to additional grants, says NIH official Joseph Ellis.

At the same time, some are concerned that wide-open access to data could lead to erroneous interpretations. Epidemiologist Bruce

Psaty of the University of Washington, Seattle, co-authored a commentary in the *Journal of the American Medical Association* last November calling NHLBI's sharing of genetic epidemiology data "a bold experiment" that "holds great promise" but whose scientific value "remains to be seen." He listed a number of worries: Some investigators who weren't familiar with how NHLBI data sets were collected have already conducted erroneous risk analyses, for example. Psaty also fears that young epidemiologists could become too dependent on data mining and not learn about designing and running studies.

Clinical data sets may be even more easily misinterpreted, says Lawrence Appel of Johns Hopkins University in Baltimore, Maryland, a co-investigator on several major blood-pressure and heart-disease trials. "It sounds good, but in reality if somebody wants to use the data from a long, complex study, you really cannot do that unless you are collaborating on the study," Appel says.

Collaborating—or at least getting feedback on the reanalysis—is the best way to go, agrees Vickers—if the original investigators are willing. If he knows them, "often it goes very smoothly." But if he doesn't, he says he's often rebuffed. He thinks journals should encourage cooperation by requiring that clinical data sets be freely available. Although many biomedical journals have a general data-sharing policy, apparently only one—the *Annals of Internal Medicine*—has an explicit requirement for technical studies. Since April 2007, it has asked all authors to include a "reproducible research" statement explaining whether they will provide the study protocol, data set, and statistical code.

But authors don't have to make a commitment as a condition for acceptance, and many are simply saying the data are not available, according to Laine. The journal hopes to give clinical researchers a gentle nudge. "We view this policy as a tiny, baby step towards changing the scientific culture," says Laine.

—JOCELYN KAISER

CONFIDENTIAL DATA REMOVED

Policies
on sharing human subjects data

Organization	Applies to ...	Effective	Policy
NIH	Grants over $500,000	2003	Provide data sharing plan or explain why not
NIH	Genome-wide association studies	January 2008	Pls "encouraged" to submit cleaned data to dbGaP
FDA	Approved drugs	September 2008	Post summary results in ClinicalTrials.gov 12 months after trial
Annals of Internal Medicine	Original research papers	April 2007	State whether protocol and data set are available

Pressure points. Clinical researchers and epidemiologists are being asked to disclose summary results and share data with colleagues. FDA now requires disclosure of summary data on approved drugs and devices, and NIH encourages the sharing of clinical and population data sets.

Women Abound in NIH Trials

A campaign that began in the 1990s to tip the balance toward female participants may have been more successful than people realize

SIX YEARS AGO, A *NEW YORK TIMES* COLUMNIST confidently stated that, before Bernadine Healy became director of the National Institutes of Health (NIH) in 1991, "women were usually excluded from clinical trials." It's a popular and tenacious view, but it's hard to find evidence for it. In fact, the current ratio of women to men in U.S.-government–funded trials is about 2-to-1.

It's true that until the 1980s, women of reproductive age were often excluded from trials, ostensibly to avoid harm to fetuses. The impression of male-dominated trials was reinforced by two large men-only heart trials launched in 1972 and 1981. In 1987, NIH formally made a commitment to include more women in research and followed in 1990 with the establishment of the NIH Office of Research on Women's Health. In 1991, NIH started the 15-year Women's Health Initiative, an intensive study of postmenopausal women.

These developments notwithstanding, many women argued that more attention was needed. In Congress, Representative Patricia Schroeder (D–CO) picked up the ball after one of her staffers called her attention to the two big male heart trials. "Because they were so big and expensive," in part, they provoked "outrage," says Adele Gilpin, a physician and lawyer at the Washington, D.C., law firm of Hunton & Williams. These pressures led to congressional passage of the 1993 NIH Revitalization Act, which further emphasized inclusion of gender considerations in research.

Despite all the hoopla, argues epidemiologist Curtis Meinert of Johns Hopkins University in Baltimore, Maryland, who has been following the issue for 2 decades, there is no evidence that women were getting short shrift. In 2000, he and Gilpin published a paper with data indicating more women than men had been in trials published

Gender and Clinical Trials*			Female only %	Male only %		
Published	Total	Male and Female %	Female only %	Male only %	F/M ratio	Unknown %
1991–1995	66,370	58.4	11.6	11.5	1.01	18.6
1996–2000	101,990	63.3	11.5	10.0	1.15	15.1
2001–2005	125,759	67.1	11.6	8.6	1.36	12.7
2006	31,114	68.9	11.7	7.4	1.58	12.0

*Counts from searching PubMed for "clinical trial," limited to publications involving human beings.

in major journals between 1985 and 1995. Meinert has since updated his figures once again (see chart, above). In an unpublished paper, he states that the ratio of female-only to male-only trials climbed in the past 10 years, especially in multicenter randomized trials for which, according to PubMed searches of English-language journals, the ratio in 2006 was 3.44-to-1. He also notes that according to statistics gathered by the Office of Research on Women's Health, of the 14.8 million people enrolled in 10,758 NIH research protocols in 2006, 64% were women.

Meinert doesn't say whether the tilt toward women is bad for science. But he argues that "general perceptions have taken priority over data" in this subject. There's now a requirement that gender must be taken into account in every grant application involving human subjects. But in many cases, sex is much less relevant than other biological factors such as age, he says. He believes the regulations pertaining to women have "served to institutionalize a perception" that is "not supported by fact [and] that erodes public trust in an important enterprise." He also contends that a bias toward females has made it "politically risky to do male-only trials."

Vivian Pinn, director of the women's health office since 1993, says she doesn't buy the notion that political concerns have influenced the sex ratio in NIH-supported trials. She acknowledges that they feature "obviously a much larger percentage of women." But she believes that's probably justified given the complexity of female reproductive issues; male-only reproductive problems are mainly confined to the prostate. Plus, she says, the numbers also reflect studies that originally were men-only that are now being repeated in women.

Reproduction aside, says Pinn, there are "fewer conditions more prevalent in men than in women." Autoimmune diseases, for example, disproportionately affect women. And although heart disease occurs much earlier in men, she's not sure there's any difference in prevalence between the sexes. Finally, notes Pinn, if you look just at studies that include both sexes, the ratio only slightly favors women—roughly 53% to 46% in 2006. Phase III trials have slightly more men than women, she says.

Is the Office of Research on Women's Health still needed at NIH? Pinn still sees a role for it, mentioning in particular research on basic biological issues such as how sex may modify particular gene functions. At any rate, she adds, "looking at percentage of money spent on males versus females is not a good way to evaluate" research. **—CONSTANCE HOLDEN**

Deadly buildup. High levels of cholesterol in the artery contribute to plaque accumulation (yellow foreground), a precursor of heart disease. But assumed links between cholesterol-control drugs, plaques, and prevention of disease are being challenged.

Cholesterol Veers Off Script

Recent drug trials have produced surprising results; along with genetics research, these findings have put in question some long-held beliefs

CHOLESTEROL NUMBERS ARE A MANTRA OF medicine, and millions of us regularly supply a vial of blood to measure this waxy substance that circulates in the bloodstream. All cells need it to survive. But it also feeds plaques in the arteries that can break open, causing a heart attack. Controlling cholesterol is gospel in cardiovascular medicine; it guides treatment and sells billions of dollars' worth of drugs. It has also been reinforced by a Hollywood-like story line: A villainous "bad" cholesterol clogs arteries, and a valiant "good" cholesterol clears them.

The cholesterol hypothesis "is like religion for some people," says Harlan Krumholz, a cardiologist at Yale University. "They've been taught it in medical school. They've been taught it forever."

But Krumholz and some others say that after many decades, the cholesterol story is turning out to be messier and more nuanced than previously believed. Scientists increasingly recognize that good and bad cholesterol, though often spoken of in the same breath, are not equally well-understood. Hundreds of studies have shown that an overabundance of bad cholesterol, known as LDL (low-density lipoprotein), is associated with heart attacks. Good cholesterol, or HDL (high-density lipoprotein), is thought to be protective, but evidence for HDL's benefit is flimsier. Some scientists are now asking whether HDL is even relevant to heart-disease risk at all. Other fundamental questions persist: Why do people with healthy cholesterol levels still suffer heart attacks? Does the mechanism by which drugs tackle cholesterol matter to health? "You ask 20 people, you get 20 opinions," says Edward Fisher, director of preventive cardiology at New York University (NYU) School of Medicine.

That uncertainty is also eroding confidence in long-practiced strategies for drug development. Twenty-one years ago, the U.S. Food and Drug Administration (FDA) approved the first of the so-called statin drugs to lower LDL levels, based on the belief that lowering LDL also prevents heart disease. That proved wildly successful. But subsequent efforts to go after heart disease with a different class of cholesterol drugs produced muddled results. In addition, many people had hoped that a complementary strategy—giving drugs to raise good cholesterol, HDL—would proffer benefits akin to those of statins. But it hasn't worked out that way. A recent clinical trial that pushed up HDL failed to protect the heart.

At the same time, new genetic studies are yielding disparate results that undermine assumptions about cholesterol. This has left scientists puzzled, with some considering whether an altogether different risk factor, inflammation, is a missing link. Next month, a study of a cholesterol-lowering drug that may also reduce inflammation will report a benefit for the heart.

The bad actor

Cholesterol was first tied to heart disease in 1910, when a German chemist found that people with atherosclerosis had a high concentration of cholesterol in their aortas. Feeding rabbits cholesterol dissolved in sunflower oil caused severe atherosclerosis, cementing the connection.

A fatty substance that's both made by the body and ingested in food, cholesterol helps build cell membranes and form hormones, as well as performing many other tasks. It doesn't dissolve in the blood but instead is carted from place to place by bulky complexes called lipoproteins. Two of these travel opposite routes: LDL transports cholesterol from the liver to other tissues, and HDL is thought to carry it from other tissues, such as the arteries, back to the liver.

For many years, scientists and physicians have recognized that high LDL raises the risk of heart attacks and that lowering it saves lives. This remains the majority view. The first LDL-lowering drug, Merck's lovastatin, was approved in 1987, heralding a new era of block-buster cholesterol drugs, most of which have reaped billions of dollars in sales every year.

FDA was so confident of the link that, in approving lovastatin and five other statins that followed, it made an unusual departure: Instead of judging efficacy based on the drug's ability to improve health or increase survival, it relied on the therapy's ability to lower LDL. That confidence proved justified: In the 1990s, a trial of 4444 patients in Scandinavia showed that another statin drug, simvastatin, reduced deaths from heart disease by 42%, a number that's held up in subsequent studies. "I have the same position now that I had many years ago," that lowering LDL saves lives, says Terje Pedersen, the cardiologist and clinical trialist at Ullevål University in Oslo who led that early study for Merck.

How low to go?

On the principle that low LDL is good, many favor driving it as far down as possible. What passes for an acceptable LDL level has steadily dropped over time, says Daniel Steinberg of the University of California, San Diego. In the 1960s, he recalls, doctors didn't worry about total cholesterol levels until they topped 280. Today, less than 200 is desirable. For LDL alone, under 100 is considered healthy (and under 130 is "near optimal"). But is that good enough? Helen Hobbs, a human geneticist at the University of Texas Southwestern Medical Center in Dallas, notes that in rural China, LDL levels hover around 60 or 70, and heart disease is about 15 times lower than in the United States.

Hunter-gatherer populations living today have LDL levels of about 70. "That would suggest that's what our species evolved to have," says Thomas Lee, a Harvard physician and the network president for Partners HealthCare System in Boston, which sets cholesterol and other guidelines for doctors.

Steinberg also wishes that treatment with statins started much earlier in life. Autopsies on young men killed in the Korean War found that even at 18 or 20 years old, they already had plaque buildup in their arteries. "We've not gone as far as we can go with

LDL," says Steinberg. "The disease starts when you're a kid," and statins "should be started as early as 30."

But coaxing LDL down to 70 or so isn't easy, and statins alone often can't achieve this. Pharmaceutical companies, on the lookout for new cholesterol strategies, especially with their statin drugs losing patent protection, in the mid-1990s renewed efforts to find novel ways to lower LDL. Schering-Plough came up with a drug, ezetimibe, which acts in the gut to prevent the body from absorbing cholesterol from food and bile. (Statins work differently, blocking the synthesis of cholesterol in the liver.) Researchers hoped that adding a new cholesterol-control mechanism would make the drug more powerful than statins alone. Like the statins, ezetimibe

sailed through FDA approval because it lowers LDL; it was marketed by Merck and Schering-Plough as Zetia and was combined with a statin to be marketed as Vytorin. In 2007, Zetia and Vytorin had combined sales of more than $5 billion.

But after nearly 4 years on the market, Vytorin unexpectedly opened a Pandora's box. In January 2008, Merck announced puzzling results from a closely watched Vytorin trial called ENHANCE, in people genetically predisposed to have very high LDL. In theory, the lower the LDL level, the better off the heart should be—but

ENHANCE suggested otherwise.

The study, published in April in *The New England Journal of Medicine* (*NEJM*), found that Vytorin performed no better than a traditional statin in slowing plaque formation in cardiac arteries, a common measure of worsening atherosclerosis. The big shock, however, was that Vytorin's middling artery protection was at odds with its superior LDL-lowering performance. The drug pulled LDL down from 210 or more without treatment to a respectable average of 141, substantially better than 192 for the statin alone. But that didn't translate to healthier arteries.

Vytorin was hammered in a second study called SEAS, which found that even though the drug lowered LDL, it did not help aortic stenosis, a disease that obstructs blood flow
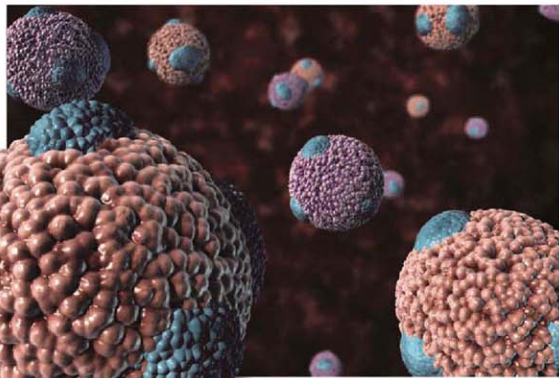
Cruising the arteries. Lipoproteins cart cholesterol around the body, but their exact functions—and how to modulate them to prevent heart disease—are under new scrutiny.

from the heart and can lead to heart failure. Results suggested it might also raise the risk of cancer. Aortic stenosis is a different disease than atherosclerosis, but there had been some research suggesting that the two were connected. SEAS, led by Pedersen, was published in September, also in *NEJM*.

Reaction to the disconnect between LDL benefits and heart health in these studies was swift and passionate. Some even questioned the link between LDL and heart disease, wondering whether lowering the first really does prevent the second. Most scientists still agree that it does and that statin drugs reduce

HDL appears to have a hand in inflammation, infection, blood clotting, oxidizing molecules, and more. "Nobody has really shown," even in animals, that HDL carries cholesterol out of arteries, as is generally believed, says Anne Tybjærg-Hansen, who studies genetics and heart disease at Copenhagen University Hospital in Denmark.

Still, animal studies have consistently shown that raising HDL prevents heart disease. Researchers, physicians, and drug companies hoped that drugs to raise HDL could become new arrows in their arsenal, making a difference to the many people who lower their LDL levels but go on to have heart attacks.

The first dramatic booster of HDL was Pfizer's drug torcetrapib. That therapy—and other HDL modulators still in development—was inspired partly by a group of families in Japan who have very high HDL and very few heart attacks. They carry mutations in a gene called cholesteryl ester transfer protein (CETP). Torcetrapib goes after this target to increase HDL levels, and by all measures it does so remarkably well. But in the same way that Vytorin lowered LDL but didn't actually help people in ENHANCE and SEAS, torcetrapib raised HDL with no measurable health benefit, and a striking downside. "That drug killed more people than it actually saved," says John Kastelein, a vascular medicine specialist at the Academic Medical Center at the University of Amsterdam in the Netherlands, who participated in the torcetrapib trial.

The 15,000-person trial found that people on torcetrapib were 60% more likely to die than those taking the statin Lipitor. In December 2006, after investing some $800 million in torcetrapib, Pfizer hastily stopped the trial, abandoned the drug, and recommended that trial participants stop taking it immediately. It's still not known why torcetrapib had the effects it did, some of which were head-scratchers, such as deaths from sepsis. More important, no one knows yet whether those effects were a fluke of this particular drug or an ominous hint that boosting HDL is dangerous.

"HDL is a good epidemiologic marker for risk," says Daniel Rader, a preventive cardiologist at the University of Pennsylvania, meaning that people with high HDL levels can rest assured that their chance of suffering heart disease is lower than average. But it's "not the best surrogate for drug development."

Genetics studies are trying to parse a link between HDL and heart disease, examining

the risk of heart disease. But the broader question—whether nonstatin drugs that lower LDL have the same effect—is now up in the air. "For me, it's an epiphany," says Steven Nissen, chair of cardiovascular medicine at the Cleveland Clinic in Ohio, who admits to having been a longtime LDL cheerleader. "It says, 'Let's be careful when you come up with a new class of drugs that lowers LDL by another mechanism.'"

Others say the results have been ridiculously overblown. "I've never seen such herd mentality in my life," especially among cardiologists and the media, says Hobbs, who is convinced that LDL is a reliable indicator of heart-disease risk. The ENHANCE work is "just a terrible paper from beginning to end." She criticizes the study for lasting only 2 years and for relying on an indirect measure of heart disease.

Hobbs points to recent genetic work in LDL that counters the clinical trial data and underscores how potent LDL can be. About 8 years ago, she set about studying the genes of people in the Dallas area with very low LDL cholesterol. She discovered that 1 in 50 African Americans has a genetic mutation that appears to lower LDL by about 30%, and heart-disease risk over 15 years by nearly 90%—far more than statins. This,

says Hobbs, is because genetics, unlike drugs begun at age 50, affects an individual over a lifetime.

Hobbs once asked the leaders of the Framingham Heart Study how many participants had an LDL of below 70 throughout life had

> "Knowing what [drugs] do to LDL and HDL doesn't tell you what they'll do to people."
> —HARLAN KRUMHOLZ, YALE

suffered a heart attack. The answer? One. "It can happen, but it's very, very rare," she says.

The HDL puzzle
If LDL is weathering some controversy, HDL is in a much deeper puddle. Data backing HDL's role in heart disease are much sparser, and HDL's function is much more poorly understood, as underscored by a recent clinical trial disaster.

Unlike LDL, which has one major receptor on the liver and one primary function,

whether people have heart-disease risks that track with HDL levels. Some have uncovered a connection, and others have not. Tybjærg-Hansen spent several years trying unsuccessfully to link low HDL to atherosclerosis in humans. An analysis of population studies in Denmark covering roughly 44,000 people, published by her group in June in *The Journal of the American Medical Association*, found no increase in heart disease among those with low HDL caused by a set of genetic mutations. Tybjærg-Hansen admits that her work gives HDL believers "high blood pressure." "I think HDL is a bystander; it's a lipid transporter in the body somehow, but I don't think it has anything to do with risk" of heart disease, she says.

> "I think HDL is a bystander; ... I don't think it has anything to do with risk."
>
> —ANNE TYBJÆRG-HANSEN, COPENHAGEN UNIVERSITY HOSPITAL

Tybjærg-Hansen blames ties between academics and drug companies that she says prevent new paradigms from bubbling up. Yale's Krumholz agrees, noting that markers in the blood may have more complex interactions with disease than thought or may not track as expected if a drug has certain side effects. "This recent group of trials should fundamentally change the way we think," he says. "Knowing what [drugs] do to LDL and HDL doesn't tell you what they'll do to people."

Companies haven't abandoned HDL drugs, however. Rader, who is working on such therapies in concert with drug companies, admits that he and the companies putting up the money are taking a risk, especially in light of torcetrapib's downfall. "But my view is, we need to test this," he says, to determine whether torcetrapib is a fluke or whether it's conveying a broader message.

A handful of large clinical trials are testing whether targeting HDL can help; results won't be available for several more years. "Our one major attempt to increase HDL failed," says Paul Ridker, a cardiologist at Brigham & Women's Hospital in Boston. It's time, he says, to look beyond cholesterol when considering how to stop atherosclerosis.

A new strategy

Ridker is focused on another factor that he thinks is involved in heart disease: c-reactive protein (CRP). High CRP has long been considered a marker of inflammation, often found in people with other risk factors, such as obesity and diabetes. But Ridker goes a step further, arguing that when CRP is lower, plaques are less likely to break open and kill their hosts. And not only that: He and a growing number of others argue that statin drugs can wrestle down CRP levels, just as they reduce LDL.

The theory that statins do more than lower cholesterol is controversial. "It just amazes me that so many people have bought into this idea that statins have all these magical properties," says Rader, who is dubious that the drugs have

CREDIT: ANNE TYBJÆRG-HANSEN

anti-inflammatory effects that prevent heart disease. A survey published in 2005 of 19 trials with more than 80,000 people found that it didn't matter how LDL was reduced; the effect on heart disease was the same.

Scientists will soon have more hard data to help them assess whether controlling inflammation with statins helps the heart. Ridker is running an unusual clinical trial named JUPITER, which offers statins to people with an LDL of under 130—considered relatively healthy—but high levels of CRP. The trial is enormous, with more than 17,000 people in 26 countries. By comparing a commonly used statin, Crestor, made by the drug company AstraZeneca, with a placebo, it is looking to see if the high-CRP group reaps a benefit. Like nearly all cholesterol drug trials, this one is funded by the drug's maker.

Results haven't been released, but AstraZeneca announced in March that it was halting the trial early because of "unequivocal evidence of benefit." Ridker will elaborate on the results at a cardiology meeting in November. An outspoken CRP proponent, he holds a patent on a method of measuring CRP in the blood. Hobbs, who disagrees with Ridker's perspective, thinks he is picking an easy target by focusing on people with LDL values of up to 130, because he "is talking about LDLs that are still high" compared with "our ancestors," she says. Pushing LDL to lower levels in this group, she believes, may explain why JUPITER succeeded: It cut down LDL, just as earlier statin trials have done.

Meanwhile, cardiologists and others are considering the clouds in their crystal ball, hoping to reach beyond decades-old strategies and more precisely predict an individual's risk of heart disease. Even the biggest proponents of the cholesterol hypothesis admit that there's more to the equation than lowering LDL. "What is causing the rest of that risk?" asks Fisher of NYU. It's the question to tackle next.

–JENNIFER COUZIN

Collective Behavior in an Early Cambrian Arthropod

Xian-Guang Hou,[1]* Derek J. Siveter,[2] Richard J. Aldridge,[3] David J. Siveter[3]

Examples that indicate collective behavior (1) in the fossil record are rare. We report chain-like associations of individuals belonging to a *Waptia*-like (crustaceanomorph) arthropod from the Chengjiang Lagerstätte, China, providing evidence that such behavior was present in the early Cambrian (about 525 million years ago), coincident with the major diversification of the Metazoa. The robustly integrated, linear configuration of these specimens is unique for any arthropod, fossil or living.

The material comprises 22 complete or partial chains and one isolated individual. The species has a folded carapace, an abdomen of at least six segments, and a telson bearing a caudal furca. The individuals are flattened, carapace compression being accommodated by peripheral crumpling and ventral overlap, apparently postmortem (right-over-left and left-over-right modes occur; Fig. 1C, carapace at left). Appendages are unknown.

Chains comprise 2 to 20 individuals. We cannot refute the hypothesis that all individuals point anteriorly, with the insertion of at least the telson and caudal furca into the carapace behind. The linkage was extremely robust: Individuals are variously stretched, bent, twisted, or telescoped together but have not parted. We interpret that the chain fell to the sea bottom, telescoping of one carapace into another occurring where the arrival angle was steep and other distortions resulting from landing under tension, directional change, or contortion. The preserved pattern shows that the individuals were not merely touching or arranged *en echelon* in life and is inconsistent with the idea that they occupied a burrow (for which evidence is lacking); rather, it indicates that the chain was pelagic.

Such a robustly integrated and strictly linear aggregation of specimens, all with the same polarity, is unique in arthropods. The trilobites in burrows and priapulid tubes (2) are not interlocked and are often less strictly aligned. In rare instances where extant arthropods show linear aggregation, they are more trains than chains, for example, the migratory queues of the lobster *Palinurus* (3) and the feeding lines of "processionary" moth caterpillars (4) and of ants (5). Among other living invertebrates, linear aggregates appear uncommon. Analogous chains occur in the thaliacean tunicates (salps), which are hermaphroditic and pelagic. In *Salpa fusiformis*, zooids maintain attachment by plaques, with zooid and chain axes parallel, thereby optimizing swimming and feeding-related migration (6).

In the case of the Chengjiang chains, aggregation for feeding is untenable because each individual would cover the oral opening of the individual behind. There is no precedent among arthropods of comparable aggregation for fertilization. Aggregation for local or wider migration may explain the Chengjiang chains, although they would represent a particularly robustly integrated example. Aggregation for defense, possibly as part of a mi-gratory strategy, might have increased the survival potential of individuals.

The robustness of the chains might suggest an inherently structural coupling. Salp chains comprise sexual individuals budded from a solitary asexual parent. In this context, the solitary Chengjiang specimen is slightly larger than all individuals in chains; also hermaphroditism occurs among recent arthropods, including some cirripedes and decapods. However, it would be radical to propose the loss of such a fundamental reproductive strategy in crustaceans during their evolutionary history. Thus, the overall evidence suggests a behavioral strategy for the Chengjiang chains, possibly associated with migration, horizontally or vertically.

References and Notes

1. D. J. T. Sumpter, *Philos. Trans. R. Soc. London Ser. B* **361**, 5 (2006).
2. B. D. E. Chatterton, R. A. Fortey, in *Advances in Trilobite Research*, I. Rábano, R. Gozalo, D. García-Bellido, Eds., vol. 9 of *Cuadernos del Museo Geominero* (Instituto Geológico y Minero de España, Madrid, 2008), pp. 73–77.
3. W. Herrnkind, *Science* **164**, 1425 (1969).
4. M. J. Scoble, *The Lepidoptera: Form, Function and Diversity* (Oxford Univ. Press, Oxford, 1992).
5. B. Hölldobler, E. O. Wilson, *The Ants* (Belknap, Cambridge, MA, 1990).
6. Q. Bone, J.-C. Braconnot, in *The Biology of Pelagic Tunicates*, Q. Bone, Ed. (Oxford Univ. Press, Oxford, 1998), pp. 5–24.
7. Thanks to National Natural Foundation (40730211) and 973 Program of China (2006CB806400); Science and Technology Department of Yunnan (2005D00022); the Royal Society; and G. Boxshall, S. de Grave, G. McGavin, and A. Parker for discussion.

[1]Yunnan Key Laboratory for Palaeobiology, Yunnan University, Kunming, Yunnan Province, People's Republic of China. [2]University Museum and Department of Earth Sciences, University of Oxford, Parks Road, Oxford, UK. [3]Department of Geology, University of Leicester, Leicester, UK.

*To whom correspondence should be addressed. E-mail: xghou@ynu.edu.cn
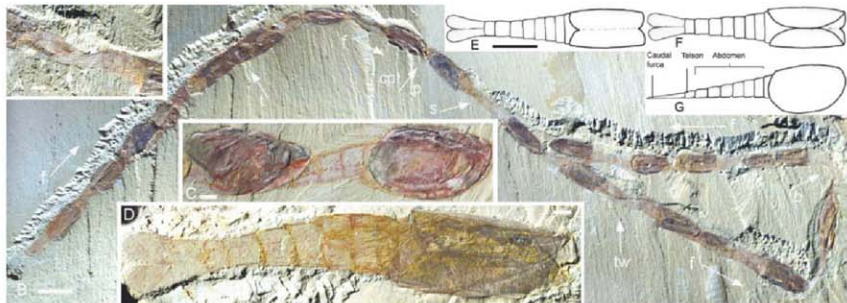
Fig. 1. *Waptia*-like arthropod, Lower Cambrian, Haikou, Yunnan. (**A**) Individual with twisted abdomen, part of chain, Yunnan Key Laboratory for Palaeobiology, YKLP 11020a. (**B**) Chain, about 20 individuals, various dorsoventral-lateral orientations, composite image (joined at cpt/p arrow), YKLP 11020a and YKLP 11020b. (**C**) Individual linked to carapace behind, lateral view, part of chain of nine individuals, YKLP 11021. (**D**) Isolated individual, subventral view, YKLP 11019. (**E** to **G**) Reconstruction shown in dorsal, ventral, and right lateral views, respectively. Scale bars in (A), (C), and (D) indicate 1 mm; in (B) and (E) to (G), 5 mm. b, s, and t indicate bent, stretched, and telescoped individuals, respectively; cpt, counterpart; f, facing direction; p, part; and tw, twisted.

The Status of the World's Land and Marine Mammals: Diversity, Threat, and Knowledge

Jan Schipper,[1,2]* Janice S. Chanson,[1,2] Federica Chiozza,[3] Neil A. Cox,[1,2] Michael Hoffmann,[1,2] Vineet Katariya,[1] John Lamoreux,[1,4] Ana S. L. Rodrigues,[5,6] Simon N. Stuart,[1,2] Helen J. Temple,[7] Jonathan Baillie,[8] Luigi Boitani,[3] Thomas E. Lacher Jr.,[2,4] Russell A. Mittermeier,[2] Andrew T. Smith,[9] Daniel Absolon,[10] John M. Aguiar,[2,4] Giovanni Amori,[11] Noura Bakkour,[2,12] Ricardo Baldi,[13,14] Richard J. Berridge,[15] Jon Bielby,[8,16] Patricia Ann Black,[17] J. Julian Blanc,[18] Thomas M. Brooks,[2,19,20] James A. Burton,[21,22] Thomas M. Butynski,[23,24] Gianluca Catullo,[25] Roselle Chapman,[26] Zoe Cokeliss,[8] Ben Collen,[8] Jim Conroy,[27] Justin C. Cooke,[28] Gustavo A. B. da Fonseca,[29,30] Andrew E. Derocher,[31] Holly T. Dublin,[32] J. W. Duckworth,[33] Louise Emmons,[34] Richard H. Emslie,[35] Marco Festa-Bianchet,[36] Matt Foster,[2] Sabrina Foster,[37] David L. Garshelis,[38] Cormack Gates,[39] Mariano Gimenez-Dixon,[40] Susana Gonzalez,[41] Jose Fernando Gonzalez-Maya,[42] Tatjana C. Good,[43] Geoffrey Hammerson,[44] Philip S. Hammond,[45] David Happold,[46] Meredith Happold,[46] John Hare,[47] Richard B. Harris,[48] Clare E. Hawkins,[49,50] Mandy Haywood,[51] Lawrence R. Heaney,[52] Simon Hedges,[33] Kristofer M. Helgen,[34] Craig Hilton-Taylor,[7] Syed Ainul Hussain,[53] Nobuo Ishii,[54] Thomas A. Jefferson,[55] Richard K. B. Jenkins,[56,57] Charlotte H. Johnston,[9] Mark Keith,[58] Jonathan Kingdon,[59] David H. Knox,[60] Kit M. Kovacs,[61,62] Penny Langhammer,[9] Kristin Leus,[63] Rebecca Lewison,[64] Gabriela Lichtenstein,[65] Lloyd F. Lowry,[66] Zoe Macavoy,[16] Georgina M. Mace,[16] David P. Mallon,[67] Monica Masi,[25] Meghan W. McKnight,[4] Rodrigo A. Medellín,[69] Patricia Medici,[70,71] Gus Mills,[72] Patricia D. Moehlman,[73] Sanjay Molur,[74,75] Arturo Mora,[76] Kristin Nowell,[77] John F. Oates,[78] Wanda Olech,[79] William R. L. Oliver,[80] Monik Oprea,[34] Bruce D. Patterson,[52] William F. Perrin,[55] Beth A. Polidoro,[1] Caroline Pollock,[7] Abigail Powel,[81] Yelizaveta Protas,[82] Paul Racey,[56] Jim Ragle,[1] Pavithra Ramani,[37] Galen Rathbun,[83] Randall R. Reeves,[84] Stephen B. Reilly,[55] John E. Reynolds III,[85] Carlo Rondinini,[3] Ruth Grace Rosell-Ambal,[86] Monica Rulli,[25] Anthony B. Rylands,[2] Simona Savini,[25] Cody J. Schank,[37] Wes Sechrest,[37] Caryn Self-Sullivan,[87] Alan Shoemaker,[88] Claudio Sillero-Zubiri,[89] Naamal De Silva,[2] David E. Smith,[37] Chelmala Srinivasulu,[90] Peter J. Stephenson,[91] Nico van Strien,[92] Bibhab Kumar Talukdar,[93] Barbara L. Taylor,[55] Rob Timmins,[94] Diego G. Tirira,[95] Marcelo F. Tognelli,[96,97] Katerina Tsytsulina,[98] Liza M. Veiga,[99] Jean-Christophe Vié,[1] Elizabeth A. Williamson,[100] Sarah A. Wyatt,[2] Yan Xie,[101] Bruce E. Young[44]

Knowledge of mammalian diversity is still surprisingly disparate, both regionally and taxonomically. Here, we present a comprehensive assessment of the conservation status and distribution of the world's mammals. Data, compiled by 1700+ experts, cover all 5487 species, including marine mammals. Global macroecological patterns are very different for land and marine species but suggest common mechanisms driving diversity and endemism across systems. Compared with land species, threat levels are higher among marine mammals, driven by different processes (accidental mortality and pollution, rather than habitat loss), and are spatially distinct (peaking in northern oceans, rather than Southeast Asia). Marine mammals are also disproportionately poorly known. These data are made freely available to support further scientific developments and conservation action.

Mammals play key roles in ecosystems (e.g., grazing, predation, and seed dispersal) and provide important benefits to humans (e.g., food, recreation, and income), yet our understanding of them is still surprisingly patchy (1). An assessment of the conservation status of all known mammals was last undertaken by the International Union for Conservation of Nature (IUCN) in 1996 (2). These IUCN Red List classifications of extinction risk (fig. S1) for mammals have been used in numerous studies, including the identification of traits associated with high extinction risk (3, 4), and prioritization of species for conservation action (5). However, the 1996 assessment was based on categories and criteria that have now been superseded, and the assessments are officially outdated for about 3300 mammals never assessed since. Previously compiled global distribution maps for terrestrial mammals (6, 7) have been used in a variety of analyses, including recommending global conservation priorities (7–9), and analyzing the coverage of protected areas (10). However, nearly 700 currently recognized species, including marine mammals, were not covered in previously published analyses.

Here, we present the results of the most comprehensive assessment to date of the conservation status and distribution of the world's mammals, covering all 5487 wild species recognized as extant since 1500. This 5-year, IUCN-led collaborative effort of more than 1700 experts in 130 countries compiled detailed information on species' taxonomy, distribution, habitats, and population status, as well as the threats to them, human use, ecology of, and conservation measures for these species. All data are freely available for consultation and downloading (11).

Diversity. Mammals occupy most of the Earth's habitats. As in previous studies (8, 12), we found that land species (i.e., terrestrial, including volant, and freshwater) have particularly high levels of species richness in the Andes and in Afromontane regions in Africa, such as the Albertine Rift. We also found high species richness in Asia, most noticeably in the Hengduan mountains of southwestern China, peninsular Malaysia, and Borneo (Fig. 1A). The ranges of many large mammals have recently contracted substantially in tropical Asia (13), so local diversity was once undoubtedly even higher. Overall, the species richness pattern for land mammals is similar to that found for birds and amphibians (12), which suggests that diversity is similarly driven by energy availability and topographic complexity (14, 15).

Marine mammals concentrate in tropical and temperate coastal platforms, as well as in offshore areas in the Tasman and Caribbean seas, east of Japan and New Zealand and west of Central America, and in the southern Indian Ocean. As with land species, marine richness seems to be associated with primary productivity: Whereas land species' richness peaks toward the equator, marine richness peaks at around 40° N and S (16) (fig. S4), corresponding to belts of high oceanic productivity (17). An interesting exception is the low species richness in the highly productive North Atlantic Ocean (17). Only one species' extinction [Sea Mink, *Neovison macrodon* (18)] and one extirpation [Gray Whale, *Eschrichtius robustus* (19)] are recorded from this region, yet evidence for many local extinctions comes from historical records of species exploitation where they no longer occur ([e.g., Harp Seal, *Phoca groenlandica*, in the Baltic Sea (20); Bowhead Whale, *Balaena mysticetus*, off Labrador (21); Walrus, *Odobenus rosmarus*, in Nova Scotia (22)]. Past human exploitation may therefore have depleted natural species richness in the North Atlantic—as it probably did with land mammals in Australia (23) and the Caribbean (24).

Phylogenetic diversity is a measure that takes account of phylogenetic relationships (and hence, evolutionary history) between taxa (16, 25) (fig. S3). It is arguably a more relevant currency of diversity and less affected by variations in taxo-

225

nomic classification than species richness. Species richness (Fig. 1A) and phylogenetic diversity (Fig. 1B) are very closely related for land species ($r^2 = 0.98$) (fig. S5), but less so in the marine environment ($r^2 = 0.73$). Disproportionately high phylogenetic diversity in the southern oceans suggests that either species here are less related than elsewhere, or that current species may in fact be poorly known complexes of multiple species, with new species awaiting discovery (consistent with the poor species knowledge in this area, see below).

The size of land species' ranges varies from a few hundred square meters (Bramble Cay Melomys, *Melomys rubicola*; Australia), to 64.7

million km^2 (Red Fox, *Vulpes vulpes*; Eurasia and North America). For marine species, ranges vary from 16,500 km^2 (Vaquita, *Phocoena sinus*; Gulf of California), to 350 million km^2 (Killer Whale, *Orcinus orca*; all oceans). Despite these extremes, most species have small ranges (fig. S6): Most land taxa occupy areas smaller than the UK, and the range of most marine mammals is smaller than one-fifth of the Indian Ocean (fig. S6).

Among land mammals, restricted-range species (those 25% of species with the smallest ranges) are concentrated on highly diverse islands (e.g., Madagascar, Sri Lanka, and Sulawesi) and tropical mountain systems (e.g., Andes,

Cameroonian, and Ethiopian Highlands) (Fig. 1C). Marine restricted-range species are almost entirely found around continental platforms, particularly in the highly productive waters off the Southern Cone of South America (Fig. 1C) (17). Both land and marine patterns of endemism are thus apparently associated with highly productive areas subject to strong environmental gradients (altitudinal in land; depth in marine).

A different perspective on patterns of species' endemism is obtained by mapping global variation in median range size (Fig. 1D). For land species, there is a strong association between landmass width and median range size: The

[1]International Union for Conservation of Nature (IUCN) Species Programme, IUCN, 28 Rue Mauverney, 1196 Gland, Switzerland. [2]Center for Applied Biodiversity Science, Conservation International, 2011 Crystal Drive, Arlington, VA 22202, USA. [3]Department of Animal and Human Biology, Sapienza Università di Roma, Viale dell'Università 32, 00185 Roma, Italy. [4]Department of Wildlife and Fisheries Sciences, Texas A&M University, College Station, TX 77843, USA. [5]Department of Zoology, University of Cambridge, Cambridge, CB2 3EJ, UK. [6]Environment and Energy Section, Mechanical Engineering Department, Instituto Superior Técnico, Portugal. [7]IUCN Species Programme, 219c Huntingdon Road, Cambridge, CB3 0DL, UK. [8]Institute of Zoology, Zoological Society of London, Regent's Park, London, NW1 4RY, UK. [9]School of Life Sciences, Arizona State University, Post Office Box 874501, Tempe, AZ 85287, USA. [10]62 Bradford Road, Combe Down, Bath, BA2 5BY, UK. [11]CNR–Institute for Ecosystem Studies, Consiglio Nazionale Delle Ricerche, Via A. Borelli, 50, 00161 Rome, Italy. [12]Department of Environmental Science and Policy, School of International and Public Affairs, Columbia University, 420 West 118th Street, New York, NY 10027, USA. [13]Centro Nacional Patagónico, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), 9120 Puerto Madryn, Argentina. [14]Patagonian and Andean Steppe Program, Wildlife Conservation Society, Boulevard Brown 2825, 9120 Puerto Madryn, Argentina. [15]Mantella-Conservation.org, 75 Plover Way, London, SE16 7TS, UK. [16]Centre for Population Biology, Imperial College London, Silwood Park, Ascot, Berkshire, SL5 7PY, UK. [17]Facultad de Ciencias Naturales e Instituto Miguel Lillo, Miguel Lillo 205, 4000 Tucumán, Argentina. [18]The Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), MIKE–Monitoring the Illegal Killing of Elephants, Post Office Box 47074, Nairobi, 00100, Kenya. [19]World Agroforestry Center (ICRAF), University of the Philippines, Los Baños, Laguna 4031, Philippines. [20]University of Tasmania, Hobart, Tasmania 7001, Australia. [21]Earthwatch Institute (Europe), 256 Banbury Road, Oxford, OX2 7DE, UK. [22]Veterinary Biomedical Sciences, The Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Edinburgh, EH25 9RG, Scotland. [23]Eastern Africa Primate Diversity and Conservation Program, Post Office Box 149, 10400 Nanyuki, Kenya. [24]Department of Bioscience and Biotechnology, Drexel University, Philadelphia, PA 19104, USA. [25]Istituto di Ecologia Applicata, via Arezzo 29, 00162 Rome, Italy. [26]23 High Street, Watlington, Oxfordshire, OX49 5PZ, UK. [27]Celtic Environment Ltd, 10 Old Mart Road, Torphins, Banchory, Kincardineshire, AB31 4JG, UK. [28]Centre for Ecosystem Management Studies, Mooshof 1a, 79297 Winden, Germany. [29]Global Environment Facility, 1818 H Street NW, G 6-602, Washington, DC 20433, USA. [30]Department of Zoology, Federal University of Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil. [31]Department of Biological Sciences, University of Alberta, Edmonton, Alberta T6G 2E9, Canada. [32]IUCN Species Survival Commission, c/o South African National Biodiversity Institute, Centre for Biodiversity Conservation, Private Bag X7, Claremont 7735, Cape Town, South Africa. [33]Wildlife Conservation Society–Asia Programs, Wildlife Conservation Society, Bronx, NY 10460, USA. [34]Department of Vertebrate Zoology, Division of Mammals, National Museum of Natural History, Smithsonian In-

stitution, Washington, DC, 20013, USA. [35]IUCN SSC (Species Survival Commission), African Rhino Specialist Group, Box 1212, Hilton 3245, South Africa. [36]Département de biologie, Université de Sherbrooke, Sherbrooke, Québec J1K 2R1, Canada. [37]Department of Environmental Sciences University of Virginia, Charlottesville, VA 22904, USA. [38]Minnesota Department of Natural Resources, 1201 East Highway 2, Grand Rapids, MN 55744, USA. [39]Environmental Science Department, University of Calgary, Calgary, AB T2N 1N4, Canada [40]Chemin de la Perroude 2 B, CH-1196 Gland, Switzerland. [41]Genética de la Conservación–IIBCE-Facultad de Ciencias, Avenida Italia 3318, Montevideo, 11600, Uruguay. [42]ProCAT International, Las Alturas, Coto Brus, Costa Rica. [43]Australian Research Council, Centre of Excellence for Coral Reef Studies, James Cook University, Townsville, Queensland 4811, Australia. [44]NatureServe, 1101 Wilson Boulevard, 15th Floor, Arlington, VA 22209, USA. [45]Sea Mammal Research Unit, Gatty Marine Laboratory, University of St. Andrews, St. Andrews, Fife, KY16 8LB, UK. [46]Australian National University, Canberra, Australian Capital Territory 0200, Australia. [47]Wild Camel Protection Foundation, School Farm, Benenden, Kent, TN17 4EU, UK. [48]Department of Ecosystem and Conservation Sciences, University of Montana, Missoula, MT 59812, USA. [49]Threatened Species Section, Department of Primary Industries & Water, General Post Office Box 44, Hobart, Tasmania 7001, Australia. [50]School of Zoology, University of Tasmania, Private Bag 5, Hobart, Tasmania 7001, Australia. [51]The Centre for Science Communication, Natural History Filmmaking, 303 Great King Street, University of Otago, Dunedin, 9054, New Zealand. [52]Department of Zoology, Field Museum of Natural History, 1400 South Lake Shore Drive, Chicago, IL 60605, USA. [53]Wildlife Institute of India, Post Box 18, Dehra Dun, 248 001, India. [54]Tokyo Woman's Christian University, 2-6-1 Zempukuji, Suginami-ku, Tokyo, 167-8585, Japan. [55]Southwest Fisheries Science Center, National Oceanic and Atmospheric Administration Fisheries, 8604 La Jolla Shores Drive, La Jolla, CA 92037, USA. [56]School of Biological Sciences, University of Aberdeen, Tillydrone Avenue, Aberdeen, AB24 2TZ, UK. [57]Madagasikara Voakajy, Boîte Postale 5181, Antananarivo (101), Madagascar. [58]Animal, Plant and Environmental Sciences, University of the Witwatersrand, Johannesburg 2050, South Africa. [59]Department of Zoology, University of Oxford, The Tinbergen Building, South Parks Road, Oxford, OX1 3PS, UK. [60]Conservation International, Private Bag X7, 7735, Claremont, South Africa. [61]Norwegian Polar Institute, 9296 Tromsø, Norway. [62]The University Centre in Svalbard, 9171 Longyearbyen, Norway. [63]Copenhagen Zoo and Conservation Breeding Specialist Group–European Regional Office, p/a Annuntiatenstraat 6, 2170 Merksem, Belgium. [64]Biology Department, San Diego State University, 5500 Campanile Road, San Diego, CA 92182, USA. [65]Instituto Nacional de Antropología y Pensamiento Latinoamericano, 3 de Febrero 1378 (1426), Buenos Aires, Argentina. [66]School of Fisheries and Ocean Sciences, University of Alaska, Fairbanks, AK 99701, USA. [67]Department of Biological Sciences, Manchester Metropolitan University, Manchester, M1 5GD, UK. [68]Nicolás Ruiz No. 100, Barrio de Guadalupe, San Cristóbal de las Casas, CP 29230, Chiapas, México. [69]Instituto de Ecología, Universidad Nacional Autónoma de México, Apartado Postal 70-275, 04510 Ciudad Universitaria, D.F., México. [70]Lowland Tapir Conservation Ini-

tiative, IPÊ–Instituto de Pesquisas Ecológicas (Institute for Ecological Research), Caixa Postal 47, Nazaré Paulista, CEP 12960-000, São Paulo, Brazil. [71]Durrell Institute of Conservation and Ecology (DICE), University of Kent, Kent, UK. [72]The Tony and Lisette Lewis Foundation, 305 Hyde Park Corner, Hyde Park, Sandton 2196, South Africa. [73]Wildlife Trust Alliance, 460 West 34th Street, New York, NY 10001, USA. [74]Zoo Outreach Organisation, Box 1683, 9A Gopal Nagar, Lal Bahadur Colony Peelamedu, Coimbatore – 641 004 Tamil Nadu, India. [75]Wildlife Information & Liaison Development Society, 9A Lal Bahadur Colony, Peelamedu, Coimbatore, Tamil Nadu 641004, India. [76]Species Unit, IUCN South America Regional Office, Calle Quiteño Libre E 15–12 y la Cumbre, Quito, Pichincha, Ecuador. [77]Cat Action Treasury, Post Office Box 332, Cape Neddick, ME 03902, USA. [78]Department of Anthropology, Hunter College of the City University of New York, New York, NY 10065, USA. [79]Warsaw University of Life Sciences, 02-787 Warsaw, Poland. [80]Philippine Biodiversity Conservation Programme, Fauna & Flora International, 3A Star Pavilion Apartments, 519 Alonso Street, Malate, Manila 1004, Philippines. [81]Victoria University of Wellington, Wellington 6140, New Zealand. [82]Department of Ecology, Evolution, and Environmental Biology, Columbia University, 10th Floor Schermerhorn Extension, MC 5557, 1200 Amsterdam Avenue, New York, NY 10027, USA. [83]Department of Ornithology and Mammalogy, California Academy of Sciences (San Francisco), Post Office Box 202, Cambria, CA 93428, USA. [84]Okapi Wildlife Associates, 27 Chandler Lane, Hudson, Quebec, J0P 1H0, Canada. [85]Mote Marine Laboratory, 1600 Ken Thompson Parkway, Sarasota, FL 34236, USA. [86]Conservation International (CI)-Philippines, 6 Maalalahanin Street, Teacher's Village, Diliman, Quezon City 1101, Philippines. [87]Sirenian International, 200 Stonewall Drive, Fredericksburg, VA 22401, USA. [88]IUCN Tapir Specialist Group, 330 Shareditch Road, Columbia, SC 29210, USA. [89]Wildlife Conservation Research Unit, Zoology Department, University of Oxford, Tubney House, Tubney OX13 5QL, UK. [90]Department of Zoology, University College of Science, Osmania University, Hyderabad, Andhra Pradesh – 500 007, India. [91]WWF International, Avenue du Mont Blanc, CH-1196 Gland, Switzerland. [92]International Rhino Foundation, c/o White Oak Conservation Center 581705 White Oak Road, Yulee, FL 32097, USA. [93]Aaranyak, 50 Samarnwoy Path (Survey), Post Office Beltola, Guwahati – 781 028, Assam, India. [94]2313 Willard Avenue, Madison, WI 53704, USA. [95]Fundación Mamíferos y Conservación, Post Office Box 17-23-055, Quito, Ecuador. [96]IADIZA (The Argentinean Arid Research Institute) – CONICET, Center for Science & Technology Mendoza, CC 507, 5500 – Mendoza, Argentina. [97]Department of Environmental Sciences & Policy, University of California, Davis, Davis, CA 95616, USA. [98]Russian Bat Research Group, Novocherkassky Prospect, 51-84, St. Petersburg 195196, Russia. [99]Zoology Department, Emilio Goeldi Museum, Avenida Perimetral, 1901, Terra Firme, Belém, Pará, Brazil. [100]Department of Psychology, University of Stirling, Stirling FK9 4LA, Scotland, UK. [101]Institute of Zoology, Chinese Academy of Science, Datuntu, Chaoyang District, Beijing, 100101, China.

*To whom correspondence should be addressed. E-mail: jan.schipper@iucn.org.
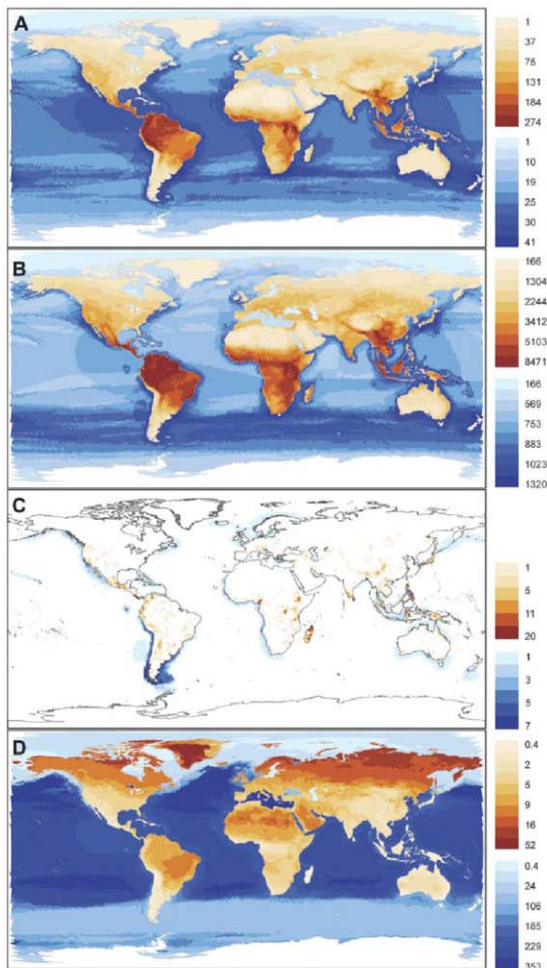
Fig. 1. Global patterns of mammalian diversity, for land (terrestrial and freshwater, brown) and marine (blue) living species, on a hexagonal grid (fig. S2). (**A**) Species richness. (**B**) Phylogenetic diversity (total branch length of the phylogenetic tree representing those species in each cell in millions of years). (**C**) Number of restricted-range species (those 25% species with the smallest ranges). (**D**) Median range size of species in each cell (in million km²).

largest ranges tend to be found across the widest part of each continent, particularly in northern Eurasia, whereas islands (e.g., in Southeast Asia and the Caribbean) and narrow continental areas (e.g., southern North and South America) tend to have narrowly distributed species. Superimposed on this general pattern, ranges also tend to be small in topographically complex areas (e.g., the Rockies, Andes, and Himalayas). These results agree with those for birds, which suggest that range sizes are constrained by the availability of land area within the climatic zones to which species are adapted (*14, 26*). Among marine species, small median range sizes are found around the continental platforms, which also suggests that steep environmental gradients (here, associated with depth) determine species' distributions. However, the global marine pattern is dominated by a latitudinal effect, with ranges generally declining toward both poles (fig. S7), which may reflect the latitudinal gradient in the overall ocean area. As with previous studies (*14, 26*), we found no support for the Rapoport rule (*27*) that the sizes of species' ranges increase with latitude.

Threat. Twenty-five percent (*n* = 1139) of all mammals for which adequate data are available (data sufficient) are threatened with extinction (Table 1). The exact threat level is unknown, as the status of 836 species having insufficient information for evaluation (Data-Deficient species) is undetermined, but is somewhere between 21% (assuming no Data-Deficient species threatened) and 36% (assuming all Data-Deficient species threatened). The conservation status of marine species is of particular concern, with an estimated 36% (range, 23 to 61%) of species threatened.

Critically Endangered species (*n* = 188) (Table 1) face a very high probability of extinction. For 29 of them, it may already be too late: Species like the Baiji (*Lipotes vexillifer*), flagged as "Possibly Extinct," have only a very small chance of still persisting (*28*). For the 76 species classified as Extinct (since 1500), no reasonable evidence suggests that they still exist. Two Extinct in the Wild species, Scimitar-horned Oryx (*Oryx dammah*) and Père David's Deer (*Elaphurus davidianus*), persist only in captivity.

Species not classified as threatened are not necessarily safe, and indeed, 323 mammals are classified as "Near Threatened" (Table 1). Many species have experienced large range and population declines in the past (e.g., Grey Wolf, *Canis lupus*, and Brown Bear, *Ursus arctos*), which are not accounted for in their current Red List status (*13*). Moreover, 52% of all species for which population trends are known are declining, including 22% of those classified as of Least Concern. These trends indicate that the overall conservation status of mammals will likely deteriorate further in the near future, unless appropriate conservation actions are put in place. On a positive note, at least 5% of currently threatened species have stable or increasing populations (e.g., European

Bison, *Bison bonasus*, and Black-footed Ferret, *Mustela nigripes*), which indicates that they are recovering from past threats.

Among land mammals, threatened species are concentrated in South and Southeast Asia (Fig. 2A). Among primates, for example, a staggering 79% (range, 76 to 80%) of species in this region are threatened with extinction. Other peaks of threat include the tropical Andes, Cameroonian Highlands, Albertine Rift, and Western Ghats in India, all regions combining high species richness (Fig. 1A), high endemism (Fig. 1, B and C), and high human pressure (29). Threatened marine species are concentrated in the North Atlantic, the North Pacific, and Southeast Asia, areas of high endemism (Fig. 1C) and high human impact (30). Low threat levels in the southern hemisphere may reflect history—as these oceans became heavily exploited much more recently—and/or knowledge gaps (see below).

Worldwide, habitat loss and degradation (affecting 40% of species assessed) and harvesting (hunting or gathering for food, medicine, fuel, and materials, which affect 17%) are by far the main threats to mammals (table S2). Yet, the relative importance of different threats varies geographically and taxonomically (Fig. 2, B and C). Among land species, habitat loss is prevalent across the tropics, which coincides with areas of high deforestation in the Americas, Africa, and Asia (Fig. 2B) (31). Harvesting is having devastating effects in Asia, but African and South American species are also affected (Fig. 2C). Harvesting affects large mammals (Cetartiodactyla, Primates, Perissodactyla, Proboscidea, and Carnivora) disproportionately: 90% in Asia, 80% in Africa, and 64% in the Neotropics (compared with 28, 15, and 11% of small mammals, respectively).

Among marine mammals, the dominant threat is accidental mortality (which affects 78% of species), particularly through fisheries by-catch and vessel strike (table S2). Although coastal areas are the most affected (Fig. 2D), accidental mortality also threatens species in off-shore waters (e.g., from purse seines in the eastern tropical Pacific). Pollution (60% of species) is the second most prevalent threat (Fig. 2E), but this designation includes a diversity of mechanisms, such as chemical contaminants, marine debris, noise, and climate change. Sound pollution (military sonar) has been implicated in mass strandings of cetaceans (32), and climate change is already impacting sea ice–dependent species (e.g., Polar Bear, *Ursus maritimus*, and Harp Seal, *Pagophilus groenlandicus*). Despite progress through international agreements, harvesting remains a major threat for marine mammals (52% of species).

Disease affects relatively few mammals (2%), but it has led to catastrophic declines in some, most dramatically the Tasmanian Devil (*Sarcophilus harrisii*) on account of facial tumor disease (33).

Threat levels are not uniform across mammalian groups (Fig. 3 and table S1). Those with disproportionately high incidence of threatened or extinct species include tapirs (Tapiridae), hippos (Hippopotamidae), tarsiers (Tarsiidae), bears (Ursidae), potoroids (Potoroidae), pigs and hogs (Suidae), and golden moles (Chrysochloridae). Families less threatened than expected include moles (Talpidae), dipodids (Dipodidae), opossums (Didelphidae), and free-tailed bats (Molossidae).

A positive association between body size and threat among mammals (3, 4), and indeed, we found that the most threatened families are dominated by large species, such as primates and ungulates, whereas the least threatened include small mammals, such as rodents and bats. Larger species tend to have lower population densities, slower life histories, and larger home ranges and are more likely to be hunted—factors that put them at greater risk (4). For smaller mammals, mostly threatened by habitat loss (34), conservation status is mainly determined by range size and location (4). However, some families of small mammals (e.g., golden moles) are also highly threatened (Fig. 3). Furthermore, although large mammals have a significantly higher fraction of threatened or extinct species ($\chi^2 \approx 0$; $P < 0.0001$), large and small mammals have suffered similar levels of extinction ($\chi^2 = 0.74$; not significant) (34).

Knowledge. Although mammals are among the best-known organisms, they are still being discovered at surprisingly high rates (1). The number of recognized species has increased by 19% since 1992 (fig. S8) and includes 349 newly described species and 512 taxa that were elevated to species level. The spatial pattern of new species description (Fig. 4A) reflects the interaction between the local state of knowledge and taxonomic effort. Peaks in Madagascar and the Amazon result from relatively high, recent taxonomic activity in these poorly known areas, whereas the lack of new species in Africa (particularly in the poorly surveyed Congo Basin) may reflect more limited efforts there.

Newly described mammals are generally poorly known (44% are Data Deficient; 13% for pre-1992 species) and disproportionately threatened [51% (29 to 73%); pre-1992 species: 23% (20 to 33%)]. These factors reinforce the concern that species may be vanishing even before they are known to science.

Newly described species have been named in areas where knowledge has increased in the recent past (Fig. 4A), whereas Data-Deficient species are concentrated in regions in need of future research (Fig. 4B). Most Data-Deficient species on land are in tropical forests (Fig. 4B), which reflects species' richness patterns (Fig. 1A); in regions where these forests are vanishing very rapidly (e.g., Atlantic Forest, West Africa, Borneo) (31), many Data-Deficient species may be dangerously close to extinction.

Marine species are less well studied than land mammals, with 38% Data Deficient (Table 1). Species that breed on land tend to be better understood, but whales, dolphins, porpoises, and sirenians are so difficult to survey that declines that should result in a Vulnerable listing would go undetected at least 70% of the time (35). Marine Data-Deficient species are particularly concentrated along the Antarctic Convergence (Fig. 4B), largely driven by the beaked whales, 19 of which are Data Deficient. A relative absence of Data-Deficient species in the northern Atlantic and Pacific Oceans may reflect higher research effort and expertise—as well as a longer history of exploitation.

Conclusion. Our results paint a bleak picture of the global status of mammals worldwide. We estimate that one in four species is threatened with extinction and that the population of one in

Table 1. Number of species in each IUCN Red List category and threat level for all mammals, and for land and marine species. Categories: EX, Extinct; EW, Extinct in the Wild; CR, Critically Endangered; EN, Endangered; VU, Vulnerable; NR, Near Threatened; LC, Least Concern; DD, Data Deficient. Threat level = [(VU+EN+CR)/(Total − DD)] × 100. The range is between [(VU+EN+CR)/Total] × 100 and [(VU+EN+CR+DD)/Total] × 100. NA, not applicable because they are not mapped.

Total and Red List category	Mammal species by habitat		
	All	Land	Marine
Number of species (% of total)			
Total	5487	5282	120
EX	76 (1.4)	NA	NA
EW	2 (0.04)	NA	NA
CR	188 (3.4)	185 (3.5)	3 (2.5)
EN	448 (8.2)	436 (8.3)	12 (10.0)
VU	505 (9.2)	497 (9.4)	12 (10.0)
NT	323 (5.9)	316 (6.0)	7 (5.8)
LC	3109 (56.7)	3071 (58.1)	40 (33.3)
DD	836 (15.2)	777 (14.7)	46 (38.3)
Threat level (%)			
Threat level	25	25	36
(range)	(21 to 36)	(21 to 36)	(23 to 61)
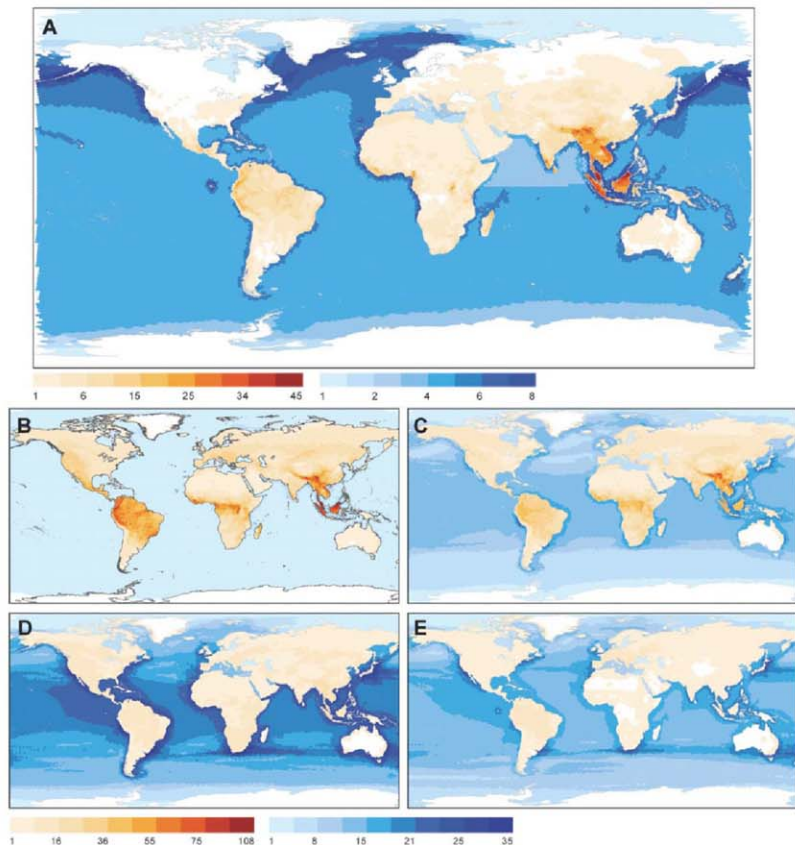
Fig. 2. Global patterns of threat, for land (brown) and marine (blue) mammals. (**A**) Number of globally threatened species (Vulnerable, Endangered or Critically Endangered). Number of species affected by: (**B**) habitat loss; (**C**) harvesting; (**D**) accidental mortality; and (**E**) pollution. Same color scale employed in (B), (C), (D) and (E) (hence, directly comparable).

two is declining. The situation is particularly serious for land mammals in Asia, through the combined effects of overharvesting and habitat loss, and for marine species, victims of our increasingly intensive use of the oceans. Yet, more than simply reporting on the depressing status of the world's mammals, these Red List data can and should be used to inform strategies for addressing this crisis (36), for example to identify priority species (5) and areas (8, 10) for con-

servation. Further, these data can be used to indicate trends in conservation status over time (37). Despite a general deterioration in the status of mammals, our data also show that species recoveries are possible through targeted conservation efforts.

References and Notes
1. D. M. Reeder, K. M. Helgen, D. E. Wilson, *Occas. Pap. Museum Texas Tech. Univ.* **269**, 1 (2007).
2. IUCN, *1996 IUCN Red List of Threatened Animals* (IUCN, Gland, Switzerland, 1996).
3. A. Purvis, J. L. Gittleman, G. Cowlishaw, G. M. Mace, *Proc. R. Soc. Lond. B. Biol. Sci.* **267**, 1947 (2000).
4. M. Cardillo *et al.*, *Science* **309**, 1239 (2005).
5. N. J. Isaac, S. T. Turvey, B. Collen, C. Waterman, J. E. M. Baillie, *PLoS One* **2**, e296 (2007).
6. W. W. Sechrest, thesis, University of Virginia, Charlottesville, VA (2003).
7. G. Ceballos, P. T. Ehrlich, J. Soberón, I. Salazar, J. P. Fay, *Science* **309**, 603 (2005).
8. G. Ceballos, P. R. Ehrlich, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 19374 (2006).

Fig. 3. Threat status of each mammalian family in relation to overall threat levels across all mammals (dashed line, 26%). Each family represented by a dot, indicating the percentage of threatened or extinct species, in relation to the total number of data sufficient species in the family. Colored bands indicate significance levels (one-tailed binomial test). Families 1 to 25 have threat levels significantly ($P <$ 0.01) different from expected (between brackets, number of threatened or extinct species/number of data sufficient species).

1 Tapiridae , Perissodactyla (4/4)
2 Hippopotamidae , Cetartiodactyla (4/4)
3 Tarsiidae , Primates (5/6)
4 Ursidae, Carnivora (6/8)
5 Potoroidae , Diprotodontia (7/11)
6 Suidae , Cetartiodactyla (10/17)
7 Chrysochloridae , Afrosoricida (10/18)
8 Hominidae , Primates (6/6)
9 Nesophontidae , Eulipotyphla (6/6)
10 Moschidae , Cetartiodactyla (7/7)
11 Hylobatidae , Primates (15/16)
12 Indriidae , Primates (12/13)
13 Capromyidae , Rodentia (17/19)
14 Lemuridae , Primates (16/19)
15 Atelidae , Primates (18/28)
16 Cercopithecidae , Primates (72/120)
17 Cervidae, Cetartiodactyla (26/45)
18 Pteropodidae , Chiroptera (69/165)
19 Bovidae, Cetartiodactyla (54/137)

18 Talpidae , Eulipotyphla (3/36)
19 Dipodidae , Rodentia (4/42)
20 Didelphidae, Didelphimorphia (9/78)
21 Molossidae , Chiroptera (11/80)
22 Phyllostomidae , Chiroptera (19/152)
23 Vespertilionidae , Chiroptera (44/325)
24 Sciuridae , Rodentia (33/238)
25 Cricetidae , Rodentia (110/612)

Fig. 4. Global patterns of knowledge, for land (terrestrial and freshwater, brown) and marine (blue) species. (**A**) Number of species newly described since 1992. (**B**) Data-Deficient species.

9. T. J. Davies et al., Proc. Natl. Acad. Sci. U.S.A. 105, 11556 (2008).
10. A. S. L. Rodrigues et al., Nature 428, 640 (2004).
11. IUCN Red List of Endangered Animals, www.iucnredlist.org/mammals.
12. R. Grenyer et al., Nature 444, 93 (2006).
13. J. C. Morrison, W. Sechrest, E. Dinerstein, D. S. Wilcove, J. F. Lamoreux, J. Mammal. 88, 1363 (2007).
14. B. A. Hawkins, J. A. F. Diniz-Filho, Glob. Ecol. Biogeogr. 15, 461 (2006).
15. R. G. Davies et al., Proc. R. Soc. Lond. B. Biol. Sci. 274, 1189 (2007).
16. Materials, methods, and additional figures and tables are available as supporting material on Science Online.
17. C. B. Field, M. J. Behrenfeld, J. T. Randerson, P. Falkowski, Science 281, 237 (1998).
18. R. Campbell, Can. Field Nat. 102, 304 (1988).
19. P. J. Bryant, J. Mammal. 76, 857 (1995).
20. J. Stora, P. G. P. Ericson, Mar. Mamm. Sci. 20, 115 (2004).
21. B. McLeod et al., Arctic 2008, 1 (2008).
22. R. R. Reeves, Atlantic Walrus (Odobenus rosmarus rosmarus): A Literature Survey and Status Report (U.S. Dept. of the Interior, Fish and Wildlife Service, Washington, DC, USA, 1978).
23. C. S. M. Turney et al., Proc. Natl. Acad. Sci. U.S.A. 105, 12150 (2008).
24. S. T. Turvey, J. R. Oliver, Y. M. Narganes-Storde, P. Rye, Biol. Lett. 3, 193 (2007).
25. D. Faith, Biol. Conserv. 61, 1 (1992).
26. C. D. L. Orme et al., PLoS Biol. 4, e208 (2006).
27. G. C. Stevens, Am. Nat. 133, 240 (1989).
28. S. T. Turvey et al., Biol. Lett. 3, 537 (2007).
29. E. Sanderson et al., Bioscience 52, 891 (2002).
30. B. S. Halpern et al., Science 319, 948 (2008).
31. F. Achard et al., Science 297, 999 (2002).
32. J. Hildebrand, in Mammal Research: Conservation Beyond Crisis, J. Reynolds III, W. F. Perrin, R. R. Reeves, S. Montgomery, T. J. Ragen, Eds. (Johns Hopkins Univ. Press, Baltimore, 2005), pp. 101–124.
33. M. Hoffmann, C. E. Hawkins, P. D. Walsh, Nature 454, 159 (2008).
34. A. Entwistle, P. Stephenson, in Priorities for the Conservation of Mammalian Biodiversity: Has the Panda Had Its Day? (Cambridge Univ. Press, Cambridge, 2000), pp. 119–139.
35. B. L. Taylor, M. Martinez, T. Gerrodette, J. Barlow, Y. N. Hrovat, Mar. Mamm. Sci. 23, 157 (2007).
36. A. S. L. Rodrigues, J. D. Pilgrim, J. F. Lamoreux, M. Hoffmann, T. M. Brooks, Trends Ecol. Evol. 21, 71 (2006).
37. S. H. Butchart et al., PLoS One 2, e140 (2007).
38. Full acknowledgments are provided in the supporting material available at Science Online.

A High Phase-Space-Density Gas of Polar Molecules

K.-K. Ni,[1]* S. Ospelkaus,[1]* M. H. G. de Miranda,[1] A. Pe'er,[1] B. Neyenhuis,[1] J. J. Zirbel,[1] S. Kotochigova,[2] P. S. Julienne,[3] D. S. Jin,[1]† J. Ye[1]†

A quantum gas of ultracold polar molecules, with long-range and anisotropic interactions, not only would enable explorations of a large class of many-body physics phenomena but also could be used for quantum information processing. We report on the creation of an ultracold dense gas of potassium-rubidium ($^{40}K^{87}Rb$) polar molecules. Using a single step of STIRAP (stimulated Raman adiabatic passage) with two-frequency laser irradiation, we coherently transfer extremely weakly bound KRb molecules to the rovibrational ground state of either the triplet or the singlet electronic ground molecular potential. The polar molecular gas has a peak density of 10^{12} per cubic centimeter and an expansion-determined translational temperature of 350 nanokelvin. The polar molecules have a permanent electric dipole moment, which we measure with Stark spectroscopy to be 0.052(2) Debye (1 Debye = 3.336×10^{-30} coulomb-meters) for the triplet rovibrational ground state and 0.566(17) Debye for the singlet rovibrational ground state.

Ultracold atomic gases have enjoyed tremendous success as model quantum systems in which one can precisely control the particles' internal degrees of freedom and external motional states. These gases make interesting many-body quantum systems when the effects of interactions between the particles, along with their quantum statistics, are important in determining the macroscopic response of the system. However, for most atomic gases the interactions are exceedingly simple: They are spatially isotropic and are sufficiently short-range to be well approximated by contact interactions. A wider range of many-body physics phenomena could be explored if the gas comprised particles with more complex interactions, such as would occur in an ultracold gas of polar molecules. Here, the electric dipole-dipole interaction is long-range and spatially anisotropic, much like the interaction of magnetic spins in condensed matter systems. Dipole-dipole interactions can be realized using atomic magnetic dipoles (1, 2) but are typically much weaker than those that could be realized for molecules with a permanent electric dipole moment. Theoretical proposals employing ultracold polar molecules range from the study of quantum phase transitions (3) and quantum gas dynamics (4) to quantum simulations of condensed matter spin systems (5) and schemes for quantum information processing (6–8).

The relative strength of dipole-dipole interactions in an ultracold gas depends critically on three parameters: the temperature T; the dipole moment, and the number density of molecules in the sample. For interaction effects to be strongly manifested, the interaction energy must be comparable to or greater than the thermal energy. This condition calls for low temperatures and large dipole moments. In addition, a high number density is needed because the dipole-dipole interaction scales as $1/R^3$, where R is the distance between particles. The combined requirements of low temperature and high density can only be met if the molecule gas has a high phase-space density, that is, the gas should be near quantum degeneracy. Recently, there has been rapid progress toward creating samples of cold polar molecules (9–13); however, it remains a challenge to create a gas where dipole-dipole interactions are observable.

Direct cooling of ground-state molecules (11–13) has thus far only attained milliKelvin final temperatures. An alternative route is to start with a high phase-space-density gas of atoms and then coherently and efficiently convert atom pairs into ground-state molecules without heat-ing the sample (14–17). To create polar molecules, the initial atomic gas must be a mixture of two types of atoms so that the resulting diatomic molecules are heteronuclear. In addition, only tightly bound molecules will have an appreciable electric dipole moment. This requirement gives rise to the considerable challenge of efficiently converting atoms that are relatively far apart into molecules of small internuclear distance, without allowing the released binding energy to heat the gas.

Preserving the high phase-space density of the initial gas while transferring atoms to deeply bound polar molecules requires coherent state transfer. Here, we report the efficient transfer of ultracold atoms into the rovibrational ground state of both the triplet and the singlet electronic ground molecular potentials and a measurement of the resulting molecules' electric dipole moments. The rovibrational ground states are clearly identified spectroscopically with precise measurements of binding energies and rotational constants, which are in excellent agreement with theory. Transfer into these states is accomplished by creating near-threshold molecules and then using a single step of stimulated Raman adiabatic passage (STIRAP) (18). Key steps in realizing efficient transfer with STIRAP are the identification of a favorable intermediate state and the ability to maintain phase coherence of the Raman lasers. With the coherent transfer to a single quantum state, we create 3×10^4 rovibrational ground-state polar molecules at a peak density of 10^{12} cm^{-3}. The molecules are created in an optical dipole trap, and their expansion energy is $k_B \times 350$ nK, where k_B is Boltzmann's constant.

The starting point for this work is a near-quantum degenerate gas mixture of fermionic ^{40}K atoms and bosonic ^{87}Rb atoms confined in an optical dipole trap. We use a magnetic-field tunable Fano-Feshbach resonance at 546.7 G

[1]JILA, National Institute of Standards and Technology (NIST) and University of Colorado, Department of Physics, University of Colorado, Boulder, CO 80309-0440, USA. [2]Physics Department, Temple University, Philadelphia, PA 19122-6082, USA. [3]Joint Quantum Institute, NIST, and University of Maryland, Gaithersburg, MD 20899-8423, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: jin@jilau1.colorado.edu (D.S.J.); ye@jila.colorado.edu (J.Y.)

Fig. 1. Diagram of the KRb electronic ground and excited molecular potentials and the vibrational levels involved in the two-photon coherent state transfer to the triplet ground state. Here, the intermediate state |e⟩ is the $v' = 10$ level of the electronically excited $2^3\Sigma$ potential. The vertical arrows are placed at the respective Condon points of the up and down transitions. The intermediate state |e⟩ has favorable transition dipole moments for both the up leg (|i⟩ to |e⟩) and the down leg (|e⟩ to |g⟩), where the initial state |i⟩ is a weakly bound Feshbach molecule and the final state |g⟩ is the rovibrational ground state ($v = 0$, $N = 0$) of the triplet electronic ground potential, $a^3\Sigma$.

(19, 20) to associate atoms into extremely weakly bound diatomic molecules (21, 22). With an adiabatic magnetic-field sweep across the resonance, we typically create a few 10^4 near-threshold molecules. The experiments described here are performed at a magnetic field of 545.9 G, where the Feshbach molecules have a binding energy of $h \times 230$ kHz (h is Planck's constant). We directly detect the Feshbach molecules using time-of-flight absorption imaging. Given the measured number, trap frequency, and expansion energy of the fermionic molecules, we find that the Feshbach molecule gas is nearly quantum degenerate with $T/T_F \approx 2$, where T_F is the Fermi temperature.

Details of the Feshbach molecule creation and detection have been described elsewhere (23, 24).

For efficient transfer from the initial state of Feshbach molecules to the final state of tightly bound ground-state molecules, the chosen intermediate, electronically excited state must have favorable wave function overlap with both the initial and final states. We first demonstrate coherent transfer of the Feshbach molecules to the rovibrational ground state of the triplet electronic ground potential, $a^3\Sigma$. This state lies about 4000 cm^{-1} above the absolute ground state, which is the rovibrational ground state of the singlet electronic ground potential, $X^1\Sigma$. Triplet and singlet
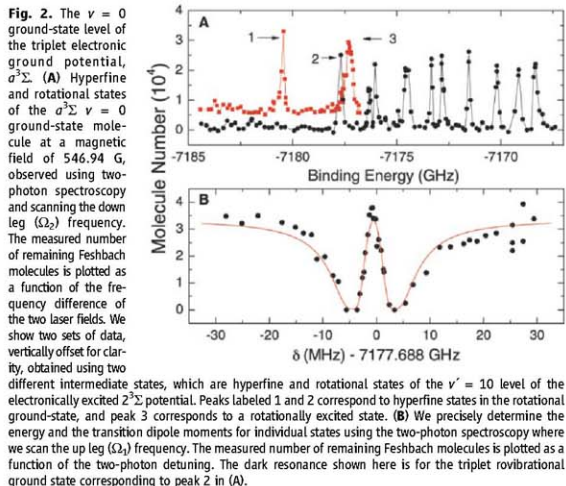
refer to a total electronic spin of the molecule that is one or zero, and our Feshbach molecules are predominantly triplet in character.

Scheme for transfer of Feshbach molecules to $a^3\Sigma$ $v = 0$ molecules. Our transfer scheme (Fig. 1) involves three molecular levels, the initial state $|i\rangle$, the intermediate state $|e\rangle$, and the final state $|g\rangle$, that are coupled by two laser fields, Ω_1 and Ω_2. The first laser field, Ω_1, drives the up transition where the wave function amplitude of the weakly bound Feshbach molecule state ($|i\rangle$) overlaps favorably near the Condon point to a deeply bound vibrational level, the 10th vibrational level, $v' = 10$, of the electronically excited $2^3\Sigma$ potential ($|e\rangle$). The Condon point is the internuclear distance where the photon energy matches the difference between the excited and ground-state potential energy curves. The second laser field, Ω_2, drives the down transition, with the Condon point near the outer turning point of the $v' = 10$ state, where its wave function overlaps strongly with the wave function for the ground vibrational level of the electronic ground $a^3\Sigma$ potential ($|g\rangle$).

The Raman system for the coherent state transfer employs a diode laser and a Ti:sapphire laser. Both lasers have tunable wavelengths around 1 μm. The two lasers are individually phase-locked to a femtosecond optical frequency comb (25), which itself is referenced to a stable 1064-nm yttrium-aluminum-garnet–Nd laser (YAG-Nd) laser. The phase-locked linewidth is measured to be narrower than 25 Hz (26, 27). The comb covers the spectral range from 532 nm to 1100 nm, with a mode spacing of 756 MHz. The large wavelength span and precision referencing capability of the comb are well suited for spectroscopy in search of previously unobserved states.

To find the intermediate state, we performed a search guided by an ab initio calculation fitted to experimental data from molecules of different isotopes, $^{39}K^{85}Rb$ (28), with the appropriate mass scalings. We located excited molecular states by measuring the loss of Feshbach molecules as a

Fig. 2. The $v = 0$ ground-state level of the triplet electronic ground potential, $a^3\Sigma$. **(A)** Hyperfine and rotational states of the $a^3\Sigma$ $v = 0$ ground-state molecule at a magnetic field of 546.94 G, observed using two-photon spectroscopy and scanning the down leg (Ω_2) frequency. The measured number of remaining Feshbach molecules is plotted as a function of the frequency difference of the two laser fields. We show two sets of data, vertically offset for clarity, obtained using two

different intermediate states, which are hyperfine and rotational states of the $v' = 10$ level of the electronically excited $2^3\Sigma$ potential. Peaks labeled 1 and 2 correspond to hyperfine states in the rotational ground-state, and peak 3 corresponds to a rotationally excited state. **(B)** We precisely determine the energy and the transition dipole moments for individual states using the two-photon spectroscopy where we scan the up leg (Ω_1) frequency. The measured number of remaining Feshbach molecules is plotted as a function of the two-photon detuning. The dark resonance shown here is for the triplet rovibrational ground state corresponding to peak 2 in (A).

Fig. 3. Time evolution of the initial state population during STIRAP state transfer and a measurement of the triplet rovibrational ground-state molecule lifetime. **(A)** Here, we monitor the Feshbach molecule population as we apply the STIRAP pulse sequence. Weakly bound Feshbach molecules are coherently transferred into the triplet rovibrational ground state after a 25-μs one-way STIRAP pulse sequence. The measured population completely disappears be-

cause the deeply bound molecules are dark to the imaging light. After a 10-μs hold, we then perform the reversed STIRAP pulse sequence that coherently transfers the ground-state molecules back to Feshbach molecules. The molecule number after the round-trip STIRAP is 1.8×10^4. Assuming equal transfer efficiency for the two STIRAP sequences, we obtain one-way

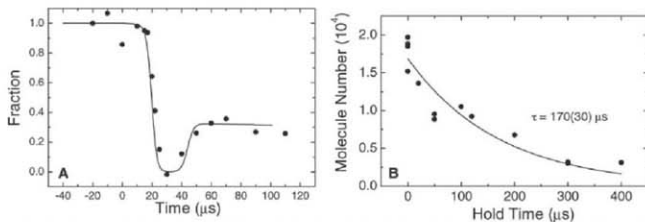transfer efficiency of 56% and an absolute number of triplet rovibrational ground-state polar molecules of 3.2×10^4 (28). **(B)** We measure the triplet $v = 0$ lifetime by varying the hold time after one-way STIRAP before transferring them back to Feshbach molecules for imaging. The lifetime is measured to be 170(30) μs.

Fig. 4. Stark spectroscopy of the triplet $v = 0$ molecules. Stark shifts of the lowest three states in the triplet $v = 0$ manifold in Fig. 2A are measured for a dc electric field in the range from 0 to 2 kV/cm. Vertical error bars represent 1 SD error from the determination of the resonance center fitted to a Lorentzian lineshape. The bottom panel shows the Stark shifts of the two lowest energy states, which are $N = 0$. Fits to the shifts of peak 1 (solid circles) and peak 2 (open circles) yield an electric dipole moment and statistical error bar of 0.052106(2) D and 0.052299(8) D respectively. The dominant systematic error in the dipole moment measurement comes from a 3% uncertainty of the electric field (horizontal error bars). A combined fit for shifts of peak 1 and peak 2 is shown as the solid curve in the bottom panel. With the electric field uncertainty, we obtain an electric dipole moment of 0.052(2) D. The top panel shows the Stark shift of peak 3 (squares) and the expected $N = 2$ curves calculated for an electric dipole moment of 0.052 D and different $|m_N\rangle$ projections.
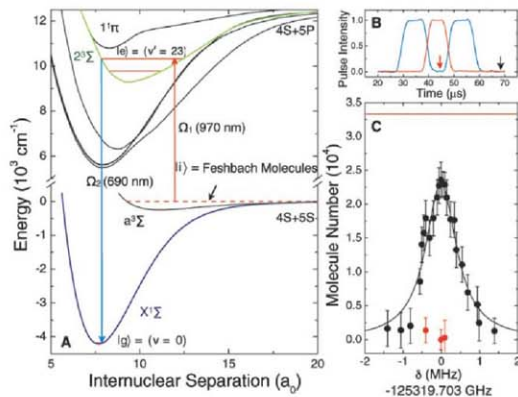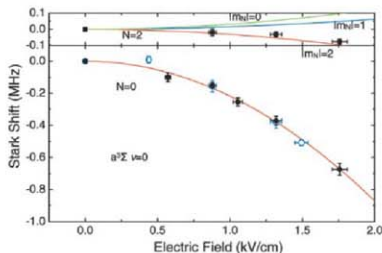
Fig. 5. Two-photon coherent state transfer from weakly bound Feshbach molecules $|i\rangle$ to the absolute molecular ground state $|g\rangle$ ($v = 0$, $N = 0$ of $X^1\Sigma$). (**A**) Transfer scheme. Here, the intermediate state $|e\rangle$ is the $v' = 23$ level of the $\Omega = 1$ component of the electronically excited $2^3\Sigma$ potential. The chosen intermediate state lies just below the $1^1\Pi$ excited electronic potential, which provides the necessary triplet-singlet spin mixing to transfer predominantly triplet character Feshbach molecules to the rovibrational ground state of the singlet electronic ground potential, $X^1\Sigma$. The vertical arrows are placed at the respective Condon points of the up and down transitions. (**B**) Normalized Raman laser intensities versus time for the round-trip STIRAP pulse sequence. We performed a 4 μs STIRAP transfer each way using a maximum Rabi frequency of $2\pi \times 7$ MHz for the downward transition (blue line) and a maximum Rabi frequency of $2\pi \times 4$ MHz for the upward transition (red line). (**C**) STIRAP lineshape. The number of Feshbach molecules returned after a round-trip STIRAP transfer is plotted as a function of the two-photon Raman laser detuning. The error bars represent the standard deviation of the measurements. The round-trip data were taken at the time indicated by the black arrow in (B). The red data points show the remaining Feshbach molecule number when only one-way transfer is performed [at the time indicated by the red arrow in (B)], where all Feshbach molecules are transferred to the ground state and are dark to the imaging light. The initial Feshbach molecule number is $3.3(4) \times 10^4$ (red solid line), and the number after round-trip STIRAP is 2.3×10^4. The round-trip efficiency is 69%, which suggests the one-way transfer efficiency is 83% and the number of absolute ground-state polar molecules is 2.7×10^4. Error bars represent 1 SD error in the molecule number.

function of the applied laser wavelength (Ω_1). This laser excites a one-photon bound-bound transition that is followed by spontaneous decay into other states. We have observed the $v' = 8$ to $v' = 12$ vibrational levels of the $2^3\Sigma$ excited potential with a roughly 1300 GHz spacing between neighboring vibrational levels, and we chose $v' = 10$ as our intermediate state. From the measured one-photon loss rate and power-broadened one-photon lineshapes, the transition dipole moment (29) is determined to be 0.004(2) ea_0 (1 $ea_0 = 2.54$ Debye $= 8.48 \times 10^{-30}$ C m).

The $a^3\Sigma$ $v = 0$ level. To search for the triplet vibrational ground state ($a^3\Sigma$; $v = 0$), we performed two-photon dark resonance spectroscopy (18) in the limit of a strong pump (Ω_2) and weak probe (Ω_1). Based on the KRb potential published by Pashov et al. (30), we calculated the triplet $v = 0$ binding energy with a predicted uncertainty of 0.1%. For the search, it was convenient to fix the frequency of the weak probe laser to resonantly drive the transition from the initial Feshbach molecule state to the $v' = 10$ intermediate state. The probe laser by itself causes complete loss of all the Feshbach molecules. We then varied the frequency of the strong coupling laser and monitored the initial state population after pulsing on both laser fields simultaneously. When the Raman condition is fulfilled, that is, the initial and final state energy splitting is matched by the two-laser frequency difference, the initial state population remains (Fig. 2A).

The measured binding energy of the triplet $v = 0$ molecules is $h \times 7.18$ THz (corresponding to 240 cm^{-1}) at 545.94 G. We find that the $v = 0$ level has rich hyperfine plus rotational structure at this magnetic field (see Fig. 2A). Because the accessible final states are influenced by selection rules, we have performed the two-photon spectroscopy using two different states of the $v' = 10$ intermediate level. In addition to the triplet $v = 0$ level, we have also observed similar ground-state hyperfine structure for the $v = 1$ and $v = 2$ levels of the $a^3\Sigma$ state, which have a vibrational energy spacing of roughly 500 GHz, consistent with our theoretical prediction.

We have identified the quantum numbers of the three lowest energy triplet $v = 0$ states seen in the two-photon spectrum. The peaks labeled 1, 2, and 3 in Fig. 2A occur at a binding energy of $h \times 7.1804180(5)$ THz, $h \times 7.1776875(5)$ THz, and $h \times 7.1772630(5)$ THz, respectively. Peak 1 corresponds to the lowest hyperfine state in the rotational ground-state ($N = 0$), peak 2 is a different hyperfine state with $N = 0$, and peak 3 is the lowest energy hyperfine state with $N = 2$, where N is the rotational quantum number. This identification is based on Hund's coupling case (b), where spin and molecular rotation are essentially decoupled and the molecular hyperfine structure can be understood from calculations using a separated atom basis with the rotational progression appearing as a constant shift to all hyperfine levels. Because of parity selection rules for optical transitions, we observe only states with an even N.

The calculated rotational constant is $B = 0.5264$ GHz, which gives a predicted splitting between the $N = 0$ and $N = 2$ levels of $6B = 3.158$ GHz.

Using a dark resonance measurement such as shown in Fig. 2B, we have measured the strength of the $|e\rangle$ to $|g\rangle$ transition. Here, we fix the down leg (Ω_2) laser frequency and scan the up leg (Ω_1) laser frequency. From the width of the dark resonance for the rovibrational triplet ground-state (peak 2), we find that we can drive the transition from $v' = 10$ to the triplet $v = 0$ state with a Rabi frequency of $2\pi \times 8$ MHz. This measurement used 60 μW of laser power focused to a beam waist of 55 μm. The transition dipole moment derived from this measurement is $0.20(2)$ ea_0, which is only one order of magnitude weaker than a typical atomic optical transition.

High phase-space-density gas of triplet ground-state polar molecules. We used peak 2 as the rovibrational ground-state target for our coherent state transfer, which is performed using the counterintuitive pulse sequence of STIRAP (18). The STIRAP beams are copropagating in order to minimize photon recoil. The measured time evolution of the initial state population during a double STIRAP pulse sequence is shown in Fig. 3A. The round-trip transfer efficiency of 31% implies a one-way transfer efficiency of 56%, which corresponds to 3×10^4 triplet $v = 0$, $N = 0$ polar molecules at a peak density of 10^{12} cm^{-3}. Our transfer technique allows us to reach a single quantum state without heating. The expansion energy of the $v = 0$ molecules is measured after transferring them back to the Feshbach molecule state. Using this expansion energy and the trap frequency measured for the Feshbach molecules, the phase-space density of the polar molecule gas corresponds to $T/T_F \approx 2.5$.

We measure the lifetime of the rovibrational ground-state molecules by varying the hold time between transferring the molecules into the $v = 0$, $N = 0$ state and bringing them back to the Feshbach molecule state for imaging. The lifetime is measured to be 170 μs as shown in Fig. 3B. This lifetime may be limited by collisions with background atoms, which can induce spin

flips and cause molecules to decay into the lower lying singlet electronic ground potential. The collisional decay could be reduced either by perfecting the removal of the remaining atoms or by starting the molecule production with atom pairs tightly confined in individual sites of an optical lattice. The short lifetime of the final state currently limits the STIRAP transfer efficiency.

We demonstrate that KRb molecules in the triplet rovibrational ground state are polar by directly measuring their electric dipole moment. The predicted dipole moment for KRb triplet rovibrational ground-state molecules is $0.05(3)$ Debye (D) (31). This is nine orders of magnitude larger than the calculated 5×10^{-11} D dipole moment of the initial Feshbach molecules and only about one order of magnitude smaller than a typical polar molecule dipole moment of 1 Debye. To measure the dipole moment, we performed dc Stark spectroscopy on the three lowest energy states observed in the two-photon spectrum (Fig. 2A). We applied a dc electric field in the range from 0 to 2 kV/cm using a pair of indium tin oxide coated transparent electric-field plates that are separated by 1.3 cm outside the glass-cell–based vacuum chamber. We measured the Stark shift using the dark resonance spectroscopy discussed above. This two-photon spectroscopy measures the energy splitting between the initial and final states, and because the initial state has a negligible dipole moment, any frequency shift of the dark resonance can be attributed to the final-state Stark shift. For these measurements, we lowered the laser powers to give a dark resonance width of 500 kHz. The measured ground-state energies versus electric field are shown in Fig. 4.

The effect of a dc electric field is to couple states of opposite parity. For the $a^3\Sigma$ $v = 0$ molecules, the opposite parity states are even-N and odd-N rotational states. The two lowest energy states, which are the rotational ground state $N = 0$, exhibit similar Stark shifts. From the measured Stark shift, we find that the molecules' electric dipole moment is $0.052(2)$ D. The Stark shift of the third energy state, corresponding to peak 3, is measured to be about 10 times smaller

than that for the peak 1 and peak 2 states. This smaller Stark shift for the $N = 2$ state is consistent with an electric dipole moment of 0.052 D.

Absolute ground-state polar molecules. We have also achieved transfer of Feshbach molecules to the rovibrational ground state of the singlet electronic potential, $X^1\Sigma$, thereby producing absolute ground-state polar molecules. Reaching the singlet rovibrational ground state requires the identification of a new suitable intermediate state and the preservation of the phase coherence of the Raman laser system over a much larger spectral difference.

A favorable scheme for transferring KRb Feshbach molecules, which have a predominantly triplet character, to the rovibrational ground state of the singlet electronic potential, $X^1\Sigma$, requires an intermediate state that has strong electronic spin-orbit coupling (9, 32) in addition to good wave function overlap with both the initial state and the target state. The intermediate state we chose (Fig. 5A) is again associated with the excited potential $2^3\Sigma$. This excited potential is split into two components labeled by $\Omega = 1$ and $\Omega = 0$, corresponding to the total electronic angular momentum projection onto the molecular internuclear axis (Ω). The higher lying vibrational levels of the $\Omega = 1$ component have strong triplet-singlet spin mixing with the nearby excited singlet potential, $1^1\Pi$ (33). This spin mixing provides the necessary coupling to transfer predominantly triplet character Feshbach molecules to the singlet rovibrational ground state. In addition, these levels guarantee a large Franck-Condon factor because the down leg Condon point is near the inner turning point of the intermediate state and coincides with the bottom of the $X^1\Sigma$ potential, whereas the up leg Condon point is near the outer turning point of the intermediate state and also overlaps favorably with the Feshbach molecule state.

For our transfer scheme (Fig. 5A), we chose $v' = 23$, $\Omega = 1$ of the electronically excited $2^3\Sigma$ potential as the intermediate state. We identified this state using the single-photon spectroscopy technique discussed in the previous section. We then proceeded with two-photon dark resonance spectroscopy to search for the rovibrational ground state of the singlet electronic ground potential. The two Raman lasers are near 970 nm and 690 nm. We measured the binding energy of the singlet rovibrational ground state ($v = 0$, $N = 0$ of $X^1\Sigma$) to be $h \times 125.319703(1)$ THz (corresponding to about 0.5 eV) at 545.88 G. This deviates by only about 400 MHz (corresponding to 0.013 cm^{-1}) from our theoretical prediction based on the potential by Pashov et al. (30). We also located the rotationally excited $N = 2$ state of the singlet vibrational ground level at a binding energy of $h \times 125.313019(1)$ THz. The energy difference of the two states, $6B = 6.6836(5)$ GHz, gives a measured rotational constant $B = 1.1139(1)$ GHz, which agrees with the predicted value of 1.1140 GHz. From the dark resonance lineshape, we extract the transition strengths for both the upward and

Fig. 6. Stark spectroscopy of the singlet $v = 0$ molecules. The bottom panel shows the Stark shift of the rovibrational ground-state of the singlet potential ($v = 0$, $N = 0$ of $X^1\Sigma$), and the top panel shows the shift of the $v = 0$, $N = 2$ state. The systematic error in the applied electric field is 3% (horizontal error bars). The level difference between $N = 0$ and $N = 2$ is 6.6836(5) GHz, which yields a rotational constant B of 1.1139(1) GHz. Given the measured B, the fit of the Stark shift (line in lower panel) gives a permanent electric dipole moment of

0.566(17) D. The theory curves for $N = 2$ for different $|m_N\rangle$ projections (lines in upper panel) are calculated using the measured B and the dipole moment derived from the $N = 0$ fit.
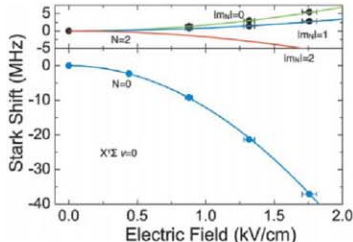
the downward transitions and obtain 0.005(2) ea_0 and 0.012(3) ea_0, respectively.

Absolute ground-state KRb polar molecules are expected to have a much larger electric dipole moment than the triplet ground-state molecules. Many theoretical efforts have calculated the dipole moment, with the results ranging from 0.5 D to 1.2 D (34). To measure the electric dipole moment, we have performed dc Stark spectroscopy on the $N = 0$ and $N = 2$ states of the singlet vibrational ground level (Fig. 6). From the measured $N = 0$ Stark shift, and the measured rotational constant B, we find that the singlet vibrational ground-state molecules have a permanent electric dipole moment of 0.566 (17) Debye, which is about 10 times as large as that of the triplet rovibrational ground state. The measured Stark shift of the $N = 2$ state is consistent with this value for the electric dipole moment. This large dipole moment allows the molecules to be polarized by modest electric fields and will facilitate exploration of interaction effects.

Finally, we demonstrate the creation of absolute ground-state molecules ($X^1\Sigma\ v = 0$, $N = 0$) using a single step of STIRAP. Using a STIRAP pulse length of 4 µs for each transfer (Fig. 5, B and C), we recovered 69% of the Feshbach molecules after a round-trip transfer. The one-way transfer efficiency, assuming equal efficiency on the two transfers, is 83%. We also performed a round-trip STIRAP pulse sequence with a 30 ms hold in between each transfer, that is, molecules are in the absolute ground state for 30 ms, and recovered 30% of the Feshbach molecules after the round-trip transfer. This shows that the absolute ground-state molecules are trapped and

have a much longer lifetime than the triplet rovibrational ground-state molecules. We measured the expansion energy of the Feshbach molecules that were transferred back after a 20 µs hold in the absolute ground state and observed no heating when comparing to the initial gas of Feshbach molecules. In the future, we anticipate that near-unity transfer efficiency should be possible with improved stabilization of the phase coherence between the two Raman lasers. This ability to create a long-lived quantum gas of ground-state polar molecules can be extended to other bialkali molecules and can pave the way for future studies of dipolar Fermi gases and dipolar Bose-Einstein condensates (35–37).

References and Notes
1. J. Stuhler et al., Phys. Rev. Lett. **95**, 150406 (2005).
2. Th. Lahaye et al., Nature **448**, 672 (2007).
3. H. P. Büchler et al., Phys. Rev. Lett. **98**, 060404 (2007).
4. L. Santos, G. V. Shlyapnikov, P. Zoller, M. Lewenstein, Phys. Rev. Lett. **85**, 1791 (2000).
5. G. Pupillo, A. Micheli, H. P. Bchler, P. Zoller, arXiv:0805.1896 (2008).
6. D. DeMille, Phys. Rev. Lett. **88**, 067901 (2002).
7. A. André et al., Nat. Phys. **2**, 636 (2006).
8. S. F. Yelin, K. Kirby, R. Cote, Phys. Rev. A **74**, 050301(R) (2006).
9. J. M. Sage, S. Sainis, T. Bergeman, D. DeMille, Phys. Rev. Lett. **94**, 203001 (2005).
10. D. Wang et al., Phys. Rev. A **75**, 032511 (2007).
11. J. D. Weinstein, R. deCarvalho, T. Guillet, B. Friedrich, J. M. Doyle, Nature **395**, 148 (1998).
12. H. L. Bethlem, G. Berden, G. Meijer, Phys. Rev. Lett. **83**, 1558 (1999).
13. B. C. Sawyer et al., Phys. Rev. Lett. **98**, 253002 (2007).
14. R. Wynar, R. S. Freeland, D. J. Han, C. Ryu, D. J. Heinzen, Science **287**, 1016 (2000).
15. K. Winkler et al., Phys. Rev. Lett. **98**, 043201 (2007).
16. S. Ospelkaus et al., Nat. Phys. **4**, 622 (2008).
17. J. G. Danzl et al., Science **321**, 1062 (2008).
18. N. V. Vitanov, M. Fleischhauer, B. W. Shore, K. Bergmann, Adv. At. Mol. Opt. Phys. **46**, 55 (2001).
19. S. Inouye et al., Phys. Rev. Lett. **93**, 183201 (2004).
20. F. Ferlaino et al., Phys. Rev. A **74**, 039903 (2006).
21. T. Köhler, K. Góral, P. S. Julienne, Rev. Mod. Phys. **78**, 1311 (2006).
22. C. Ospelkaus et al., Phys. Rev. Lett. **97**, 120402 (2006).
23. J. J. Zirbel et al., Phys. Rev. Lett. **100**, 143201 (2008).
24. J. J. Zirbel et al., Phys. Rev. A **78**, 013416 (2008).
25. S. T. Cundiff, J. Ye, Rev. Mod. Phys. **75**, 325 (2003).
26. S. M. Foreman et al., Phys. Rev. Lett. **99**, 153601 (2007).
27. When the YAG is stabilized to an ultrastable laser, such as a strontium clock, the two-photon linewidth is narrower than 1 Hz.
28. C. Amiot, J. Mol. Spec. **203**, 126 (2000).
29. The transition dipole moment given here includes contributions from both the electronic transition dipole moment and the Franck-Condon factor.
30. A. Pashov et al., Phys. Rev. A **76**, 022511 (2007).
31. S. Kotochigova, P. S. Julienne, E. Tiesinga, Phys. Rev. A **68**, 022501 (2003).
32. W. C. Stwalley, Eur. Phys. J. D **31**, 221 (2004).
33. T. Bergeman, A. J. Kerman, J. Sage, S. Sainis, D. DeMille, Eur. Phys. J. D **31**, 179 (2004).
34. M. Aymar, O. Dulieu, J. Chem. Phys. **122**, 204302 (2005).
35. C. Ticknor, Phys. Rev. Lett. **100**, 133202 (2008).
36. S. Ronen, J. L. Bohn, Phys. Rev. A **76**, 043607 (2007).
37. M. A. Baranov, Phys. Rep. **464**, 71 (2008).
38. This work has been supported by NSF, NIST, the Air Force Office of Scientific Research, and the W. M. Keck Foundation. We thank D. Wang for experimental assistance, J. Bohn for discussion, and H. J. Kimble and P. Zoller for reading the manuscript. K.-K.N. and B.N. acknowledge support from NSF, S.O. from the Alexander von Humboldt Foundation, M.H.G.de.M. from the CAPES/Fulbright, and P.S.J. from the Office of Naval Research.

28 July 2008; accepted 8 September 2008
Published online 18 September 2008;
10.1126/science.1163861
Include this information when citing this paper.

Cavity Optomechanics with a Bose-Einstein Condensate

Ferdinand Brennecke,* Stephan Ritter,* Tobias Donner, Tilman Esslinger†

Cavity optomechanics studies the coupling between a mechanical oscillator and the electromagnetic field in a cavity. We report on a cavity optomechanical system in which a collective density excitation of a Bose-Einstein condensate serves as the mechanical oscillator coupled to the cavity field. A few photons inside the ultrahigh-finesse cavity trigger strongly driven back-action dynamics, in quantitative agreement with a cavity optomechanical model. We approach the strong coupling regime of cavity optomechanics, where a single excitation of the mechanical oscillator substantially influences the cavity field. The results open up new directions for investigating mechanical oscillators in the quantum regime and the border between classical and quantum physics.

Cavity optomechanics has played a vital role in the conceptual exploration of the boundaries between classical and quantum mechanical systems (1). Fundamental questions related to this regime have found renewed interest through experimental progress with microengineered mechanical oscillators. Laser cooling of the mechanical mode (2–7) has been a substantial step toward the quantum regime (8, 9).

In general, light affects the motional degrees of freedom of a mechanical system through the radiation pressure force, which is caused by the exchange of momentum between light and matter. In cavity optomechanics, the radiation pressure-induced interaction between a single mode of an optical cavity and a mechanical oscillator is mediated by the optical path length of the cavity, which depends on the displacement of the mechanical oscillator.

New possibilities for cavity optomechanics are now emerging in atomic physics by combining the tools of cavity quantum electrodynamics (QED) (10, 11) with those of ultracold gases. Placing an ensemble of atoms inside a high-finesse cavity enhances the atom-light interaction because the atoms collectively couple to the same

Institute for Quantum Electronics, ETH Zürich, CH-8093 Zürich, Switzerland.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: esslinger@phys.ethz.ch

light mode (12–17). In the dispersive regime, this provides a large optomechanical coupling strength, tying the atomic motion to the evolution of the cavity field. Recently, a thermal gas prepared in a stack of nearly two-dimensional trapping potentials was shown to couple to the cavity field by a collective center-of-mass mode, leading to Kerr nonlinearity at low photon numbers (15) and back-action heating induced by quantum force fluctuations (18).

One goal for cavity optomechanical systems is the preparation of the mechanical oscillator in its ground state, with no thermally activated excitations present, yet at the same time providing strong coupling to the light field. We use a Bose-Einstein condensate (BEC) as the ground state of a mechanical oscillator and thereby suppress ther-

mal excitations of the oscillator. The cavity field couples to a collective density excitation of the BEC that matches the cavity mode, resulting in a large coupling strength. Despite the absence of an external restoring force for the mechanical oscillator, the framework of cavity optomechanics can be applied because only a single excitation mode of the BEC is involved (Fig. 1).

In our experimental setup (17, 19), a BEC of typically 1.2×10^5 ^{87}Rb atoms in the $|F=1, m_F=-1\rangle$ ground state is coupled to the field of an optical ultrahigh-finesse Fabry-Perot cavity, where F is the total angular momentum and m_F is the magnetic quantum number. The system is in the strong coupling regime of cavity QED; that is, the maximum coupling strength between a single atom and a single intracavity photon $g_0 = 2\pi \times$

10.9 MHz is larger than both the amplitude decay rate of the atomic excited state, $\gamma = 2\pi \times 3.0$ MHz, and that of the intracavity field, $\kappa = 2\pi \times 1.3$ MHz. Here, a circularly polarized intracavity photon couples the ground state to the $|F'=2, m_{F'}=-2\rangle$ excited state. Trapping the condensate within the cavity is accomplished by a crossed-beam dipole trap with trap frequencies $(\omega_x, \omega_y, \omega_z) = 2\pi \times$ (222, 37, 210) Hz, where x denotes the cavity axis and z the vertical axis. The cavity has a length of 178 μm and its fundamental transverse mode has a waist of 25 μm. The mode maximally overlaps with the condensate having Thomas-Fermi radii of $(R_x, R_y, R_z) = (3.3, 20.0, 3.5)$ μm, as deduced from a mean-field approximation. All experiments presented here were performed without active stabilization of the cavity length, which would give rise to an additional standing-wave potential for the atoms (15, 17, 18).

The coupled dynamics of the BEC and the cavity field is driven by continuously applying a weak pump laser field along the cavity axis (Fig. 1). The light transmitted through the cavity is monitored with a single-photon counter and serves as a probe for the dynamics of the system. With a detuning of $\Delta_a = \omega_p - \omega_a \geq 10^4\gamma$ between pump laser frequency and atomic transition frequency ω_a, spontaneous emission can be mostly neglected.

Figure 2 shows the response of the system while scanning the pump frequency across the optical resonance. The transmission signal (Fig. 2A) exhibits a sharp rising edge and subsequently regular and fully modulated oscillations (Fig. 2B) lasting for about 2.5 ms. The frequency at the start of these oscillations is about 37 kHz. It slightly decreases over the train of oscillations and does not depend on the speed at which the pump frequency is varied. Similar responses of the system were measured for lower pump strengths at the same detuning (Fig. 2C) as well as for atom-pump detunings of up to $\Delta_a = 2\pi \times$ 300 GHz, provided the pump rate was increased sufficiently. Moreover, when continuing the pump frequency scan, we observed a further train of oscillations appearing in the vicinity of a second resonance driven by circularly polarized photons of opposite handedness (17). This is in accordance with the observation that the condensate remained intact during probing, as directly inferred from absorption images taken after probing. The observed oscillatory behavior is obviously in strong contrast to a Lorentzian-shaped resonance curve, which would be expected for an atomic ensemble frozen inside the cavity (i.e., when the atomic external degree of freedom is neglected).

To describe the driven BEC-cavity dynamics, we consider a one-dimensional model in which the atomic motion along the cavity axis is quantized. We adiabatically eliminate the internal state dynamics of the atoms, as justified by the large detuning between pump laser frequency and atomic resonance. Denoting the creation operator for cavity photons by $\hat{a}^\dagger$ and the condensate wave function (normalized to the atom number N) by ψ,

Fig. 1. (A) Cavity optomechanical model system. A mechanical oscillator, here one of the cavity mirrors, is coupled via radiation pressure to the field of a cavity whose length depends on the oscillator displacement. **(B)** Coupling a BEC dispersively to the field of an optical high-finesse cavity constitutes an equivalent system. Here, a collective density excitation of the condensate acts as the mechanical oscillator, which strongly couples to the cavity field. Feedback

on the cavity field is accomplished by the dependence of the optical path length on the atomic density distribution within the spatially periodic cavity mode structure. In contrast to optomechanical systems presented so far, this mechanical oscillator is not based on the presence of an external harmonic potential (e.g., a spring); rather, it is provided by kinetic evolution of the condensate density excitation.
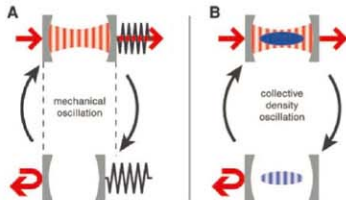
Fig. 2. (A to C) Response of the continuously driven BEC-cavity system. Shown is a single trace of the cavity transmission (averaged over 2 μs) while scanning the cavity-pump detuning at a rate of $+2\pi \times 2.9$ MHz/ms across its σ^- resonance (17). The pump rate corresponds to a mean intracavity photon number on resonance of 7.3 ± 1.8 [(A) and (B)] and 1.5 ± 0.4 (C). The photon count rate for one mean intracavity photon is 0.8 ± 0.2 MHz. The dead time of the single-photon counter is 50 ns, which leads to a saturation of high photon count rates. The pump laser was bluedetuned by $\Delta_a = 2\pi \times 32$ GHz with respect to the atomic resonance.
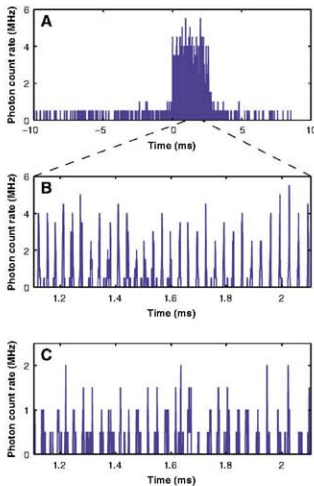
the equations of motion for the coupled system read

$$i\hbar\dot{\psi}(x) = \left[\frac{-\hbar^2}{2m}\frac{d^2}{dx^2} + \langle \hat{a}^\dagger \hat{a}\rangle \hbar U_0 \cos^2(kx) + \right.$$

$$\left. V_{\text{ext}}(x) + g_{1D}|\psi|^2 \right] \psi(x) \qquad (1)$$

$$i\dot{\hat{a}} = -\left[\Delta_c - U_0\langle\cos^2(kx)\rangle + i\kappa \right]\hat{a} + i\eta \qquad (2)$$

(20). Here, V_{ext} denotes the weak external trapping potential for atoms with mass m, and g_{1D} is the effective atom-atom interaction strength integrated along the transverse directions.

Equation 1 describes the condensate dynamics in a dynamic lattice potential. Its depth is determined by the mean intracavity photon number $\langle\hat{a}^\dagger\hat{a}\rangle$, which depends in a nonlocal and nonlinear way on the condensate wave function ψ itself. For a single intracavity photon, the potential depth is given by the light shift $U_0 = g_0^2/\Delta_a$. The coupling between the cavity field and the atomic external degrees of freedom is mediated by the spatial overlap $\langle\cos^2(kx)\rangle = \int|\psi(x)|^2 \cos^2(kx)dx$ between atomic density and cavity mode structure, with wavelength $\lambda = 2\pi/k = 780$ nm. This mode overlap determines the effective refractive index of the condensate, and with it the frequency shift of the empty cavity resonance in Eq. 2. The pump laser, which coherently drives the cavity field at a rate η, is detuned from the empty cavity frequency ω_c by $\Delta_c = \omega_p - \omega_c$.

The observed BEC-cavity dynamics (Fig. 2) can be described in a homogeneous two-mode model where the macroscopically occupied zero-momentum state is coupled to the symmetric superposition of the $\pm 2\hbar k$ momentum states via absorption and stimulated emission of cavity photons. The corresponding wave function reads $\psi(x, t) = c_0(t) + c_2(t)\sqrt{2}\cos(2kx)$, with probability amplitudes c_0 and c_2 fulfilling $|c_0(t)|^2 + |c_2(t)|^2 = N$. The mode overlap is then given by $\langle\cos^2(kx)\rangle = [N + \sqrt{2}\,\text{Re}(c_0^*c_2)]/2$. It oscillates under kinetic evolution of ψ at 4 times the recoil frequency $\omega_{\text{rec}} = \hbar k^2/(2m) = 2\pi \times 3.8$ kHz, with the atom-atom interactions being neglected at this stage. This leads to the natural definition of a harmonic oscillator with displacement $X = 2\sqrt{1/N}\,\text{Re}(c_0^*c_2)$ in units of the oscillator length, and its conjugate variable $P = \hbar\sqrt{1/N}\,\text{Im}(c_0^*c_2)$. The equations of motion (Eqs. 1 and 2) for $|c_2|^2/|c_0|^2 \ll 1$ then read as

$$\ddot{X} + (4\omega_{\text{rec}})^2 X = -\omega_{\text{rec}}U_0\sqrt{8N}\langle\hat{a}^\dagger\hat{a}\rangle \qquad (3)$$

$$i\dot{\hat{a}} = -(\Delta + i\kappa)\hat{a} + i\eta \qquad (4)$$

and describe a mechanical oscillator coupled via the radiation pressure force to the field of a cavity whose resonance frequency shift $\Delta = \Delta_c - U_0N/2 - U_0/2\sqrt{N/2}\,X$ depends linearly on the oscillator displacement X (21). The coupling strength between the optical and mechanical resonator can be varied via the atom-pump detuning Δ_a,

which allows us to experimentally enter the regime of strong coupling.

From this equivalence to cavity optomechanics, we can anticipate bistable behavior. Indeed, for pump rates larger than a critical value η_{cr}, we find three steady-state solutions for the oscillator displacement X, with two of them being stable (Fig. 3A) (15, 22, 23). The system prepared below the resonance will follow the steady-state branch until reaching the lower turning point, where a non–steady-state dynamics is excited. This dynam-

ics is governed by the time scale of the mechanical motion because the cavity damping is two orders of magnitude faster. Thus, we can assume that the cavity field follows the mechanical motion adiabatically and that retardation effects, underlying cooling and amplification, are negligible (21). Numerical integration of the coupled Eq. 3 for our experimental parameters results in fully modulated oscillations of the cavity field and cavity output (Fig. 3B), which is in very good agreement with the experimental observations (Fig. 2B).
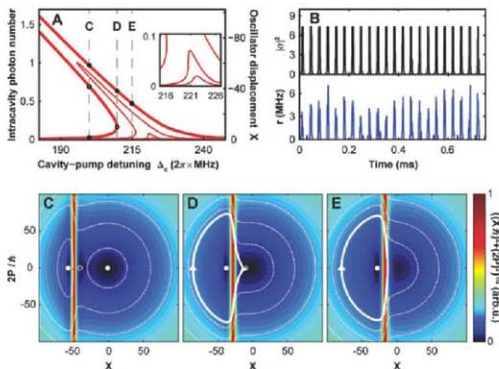
Fig. 3. Steady-state and dynamical behavior of the BEC-cavity system in the two-mode model. (**A**) Mean intracavity photon number and corresponding oscillator displacement X versus cavity-pump detuning Δ_c for the steady-state solutions of Eqs. 3 and 4. The curves correspond to mean intracavity photon numbers on resonance of $\eta^2/\kappa^2 = 0.02$, 0.07, 1, and 7.3, and a pump-atom detuning of $\Delta_a = 2\pi \times 32$ GHz. The inset highlights the bistable behavior for pump amplitudes larger than $\eta_{cr} \approx 0.27\kappa$. (**B**) Intracavity photon number $|\alpha|^2$ and corresponding transmission count rate r (including detection shot noise and averaging over 2 μs) for the system circling along the solid line in (E). For integration of the equations of motion, a coherent intracavity field α was assumed. (**C** to **E**) Evolution of the system in the mechanical phase space depicted for three subsequent situations $[\Delta_c = 2\pi \times (200, 209.7, 215)$ MHz] corresponding to the markers in (A) and $\eta^2/\kappa^2 = 7.3$. The stable and unstable steady-state configurations are displayed as solid and open circles, respectively. Dashed lines show representative evolutions for different starting conditions. Coloring indicates the modulus of the time evolution field $(\dot{X}, 2\dot{P}/\hbar)$. The solid lines in (D) and (E) correspond to the experimental situation in Fig. 2A and show the evolution of the system while scanning Δ_c at a rate of $2\pi \times 2.9$ MHz/ms across the resonance with the system initially prepared in the lower stable solution.

Fig. 4. Oscillation frequency while scanning over the resonance. The frequency within time bins of 50 μs was obtained from a peak-detection routine applied to the cavity transmission data averaged over 10 μs. The data (solid circles) are an average, ±SE, over 23 traces referenced to the start of the oscillations. Open circles show the result of a numerical integration of the one-dimensional

system (Eqs. 1 and 2) taking atomic interactions and external trapping into account (26). The mean intracavity photon number on resonance was 3.6 ± 0.9. To fit the slope of the data, we added the effect of a dynamically induced atom loss during the time of oscillations of 1.5×10^3/ms to the experimental frequency chirp of $\dot{\Delta}_c = 2.9$ MHz/ms. The background rate of atom loss was measured to be 45/ms, and an atom number of $(116 \pm 18) \times 10^3$ was deduced from absorption images taken after the oscillations.
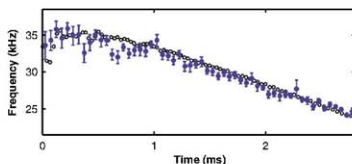
Further insight is gained by examining the dynamics in the phase space of the mechanical oscillator, spanned by X and P (Fig. 3, C to E). Without a cavity field, the time evolution would simply correspond to a clockwise rotation at $4\omega_{rec}$. Yet when photons enter the cavity, the evolution is affected by light forces. This is the case along the vertical resonance line determined by the resonance condition $\Delta = 0$, as shown in Fig. 3, C to E (red line).

Initially, the condensed atoms are prepared at the stable phase-space point $(X, P) = 0$ (Fig. 3C). Increasing the detuning Δ_c across the resonance renders the system unstable and triggers parametrically excited oscillations, as indicated by the solid line in Fig. 3D. The evolution along this path is dominated by the free oscillator dynamics, which are periodically interrupted by the interaction with the cavity light field (Fig. 3, D and E). This behavior is closely related to the matter-wave dynamics of a kicked rotor that is operated at an antiresonance where the accumulated phase factor between two kicks inhibits occupation of higher-momentum modes (24).

The frequency of these oscillations decreases continuously over observation time (Fig. 4). This is expected when actively scanning the cavity-pump detuning Δ_c, which shifts the resonance line in the phase-space diagram and leads to an adiabatic change of the system's circling path (compare Fig. 3, D and E).

A precise quantitative understanding of the observed frequency and its decrease is obtained when taking into account atom-atom interactions, the external trapping potential, and atom losses. The atom-atom interactions result in a shift of the bare oscillation frequency $4\omega_{rec} = 2\pi \times 15.1$ kHz by the mean field energy, which in the Thomas-Fermi limit equals 4/7 times the chemical potential $\mu = 2\pi \times 2.4$ kHz (25). The trapping potential gives rise to a Fourier-limited broadening of the initial momentum distribution and accordingly introduces a damping of the free-running oscillator dynamics. This suppresses a double-peak structure in the transmitted light, which would be expected at the onset of oscillations for the homogeneous two-mode model (see Fig. 3D). An enhanced atom loss during the oscillations accelerates the observed frequency shift by a factor of 2. The numerical integration of the full one-dimensional model (Eqs. 1 and 2) yields very good agreement with our data (Fig. 4).

The quantitative agreement between experiment and semi-classical theory, together with the observation of very narrow peaks in the fully modulated cavity transmission, indicates that our system is well localized in the phase space of the mechanical oscillator. Using a second quantized picture where the BEC acts as the vacuum state of the mechanical oscillator mode, we have estimated the expectation value for thermal excitations in this mode. It is found to be below 0.01 for a realistic condensate fraction of 90% (26). This extremely pure preparation of the ground state of a mesoscopic mechanical oscillator is

possible because the cavity couples only to one specific excitation mode. Because of the high finesse of the cavity, a single coherent mechanical excitation leads to a detectable shift of the cavity resonance by 0.7κ. Entering this strongly coupled quantum regime of cavity optomechanics promises to be ideal for testing fundamental questions of quantum mechanics (8, 9).

From the perspective of quantum many-body physics, we have investigated a Bose gas with weak local interactions subject to nonlocal interactions mediated by the cavity field. Experimentally, it should also be possible to enter the strongly correlated regime where local interactions dominate over the kinetic energy. In this case, the nonlocal coupling is predicted to give rise to novel quantum phases (27–29).

References and Notes
1. V. B. Braginsky, Y. I. Vorontsov, K. S. Thorne, *Science* **209**, 547 (1980).
2. C. Höhberger Metzger, K. Karrai, *Nature* **432**, 1002 (2004).
3. A. Schliesser, P. Del'Haye, N. Nooshi, K. J. Vahala, T. J. Kippenberg, *Phys. Rev. Lett.* **97**, 243905 (2006).
4. O. Arcizet, P.-F. Cohadon, T. Briant, M. Pinard, A. Heidmann, *Nature* **444**, 71 (2006).
5. S. Gigan *et al.*, *Nature* **444**, 67 (2006).
6. T. Corbitt *et al.*, *Phys. Rev. Lett.* **98**, 150802 (2007).
7. J. D. Thompson *et al.*, *Nature* **452**, 72 (2008).
8. S. Mancini, V. I. Man'ko, P. Tombesi, *Phys. Rev. A* **55**, 3042 (1997).
9. W. Marshall, C. Simon, R. Penrose, D. Bouwmeester, *Phys. Rev. Lett.* **91**, 130401 (2003).
10. C. J. Hood, T. W. Lynn, A. C. Doherty, A. S. Parkins, H. J. Kimble, *Science* **287**, 1447 (2000).
11. P. W. H. Pinkse, T. Fischer, P. Maunz, G. Rempe, *Nature* **404**, 365 (2000).
12. B. Nagorny, Th. Elsässer, A. Hemmerich, *Phys. Rev. Lett.* **91**, 153003 (2003).
13. A. T. Black, H. W. Chan, V. Vuletić, *Phys. Rev. Lett.* **91**, 203001 (2003).
14. S. Slama, S. Bux, G. Krenz, C. Zimmermann, P. W. Courteille, *Phys. Rev. Lett.* **98**, 053603 (2007).
15. S. Gupta, K. L. Moore, K. W. Murch, D. M. Stamper-Kurn, *Phys. Rev. Lett.* **99**, 213601 (2007).
16. Y. Colombe *et al.*, *Nature* **450**, 272 (2007).
17. F. Brennecke *et al.*, *Nature* **450**, 268 (2007).
18. K. W. Murch, K. L. Moore, S. Gupta, D. M. Stamper-Kurn, *Nat. Phys.* **4**, 561 (2008).
19. A. Öttl, S. Ritter, M. Köhl, T. Esslinger, *Rev. Sci. Instrum.* **77**, 063118 (2006).
20. P. Horak, S. M. Barnett, H. Ritsch, *Phys. Rev. A* **61**, 033609 (2000).
21. T. J. Kippenberg, K. J. Vahala, *Opt. Express* **15**, 17172 (2007).
22. P. Meystre, E. M. Wright, J. D. McCullen, E. Vignes, *J. Opt. Soc. Am. B* **2**, 1830 (1985).
23. A. Dorsel, J. D. McCullen, P. Meystre, E. Vignes, H. Walther, *Phys. Rev. Lett.* **51**, 1550 (1983).
24. F. L. Moore, J. C. Robinson, C. F. Bharucha, B. Sundaram, M. G. Raizen, *Phys. Rev. Lett.* **75**, 4598 (1995).
25. J. Stenger *et al.*, *Phys. Rev. Lett.* **82**, 4569 (1999). Erratum: *Phys. Rev. Lett.* **84**, 2283(E) (2000).
26. See supporting material on *Science* Online.
27. C. Maschler, H. Ritsch, *Phys. Rev. Lett.* **95**, 260401 (2005).
28. J. Larson, B. Damski, G. Morigi, M. Lewenstein, *Phys. Rev. Lett.* **100**, 050401 (2008).
29. D. Nagy, G. Szirmai, P. Domokos, *Eur. Phys. J. D* **48**, 127 (2008).
30. We thank K. Baumann, P. Domokos, C. Guerlin, I. Mekhov, H. Ritsch, and A. Vukics for stimulating discussions. Supported by the SCALA Integrated Project (European Union) and QSIT (ETH Zürich).

Carbon Nanotube Arrays with Strong Shear Binding-On and Easy Normal Lifting-Off

Liangti Qu,[1] Liming Dai,[1]* Morley Stone,[2] Zhenhai Xia,[3] Zhong Lin Wang[4]*

The ability of gecko lizards to adhere to a vertical solid surface comes from their remarkable feet with aligned microscopic elastic hairs. By using carbon nanotube arrays that are dominated by a straight body segment but with curly entangled top, we have created gecko-foot–mimetic dry adhesives that show macroscopic adhesive forces of ~100 newtons per square centimeter, almost 10 times that of a gecko foot, and a much stronger shear adhesion force than the normal adhesion force, to ensure strong binding along the shear direction and easy lifting in the normal direction. This anisotropic force distribution is due to the shear-induced alignments of the curly segments of the nanotubes. The mimetic adhesives can be alternatively binding-on and lifting-off over various substrates for simulating the walking of a living gecko.

The unusual ability of gecko lizards to climb on any vertical surface and hang from a ceiling with one toe has inspired scientific interest for decades. Only in the past few years has progress been made in understanding the mechanism that allows the gecko to defy gravity in climbing vertical surfaces (1, 2). Recent studies revealed the remarkable gecko foot with countless specialized keratinous aligned

microscopic elastic hairs (3 to 130 μm in length), called setae, splitting into even smaller spatulae (0.2 to 0.5 μm in diameter) at the end (1, 2). It is these spatulae that come in close contact with the surface to induce strong van der Waals (vdW) forces (~10 N cm^{-2}) (1–3) to hold gecko lizards onto a vertical wall. Attempts have been made to mimic gecko feet by using microfabricated arrays of polymer pillars (4, 5), but the polymeric dry

adhesives have produced a maximum adhesive force of ~3 N cm^{-2}, about one-third of the value obtained by geckos.

Having an extraordinary high aspect ratio and exceptional mechanical strength (6), vertically aligned carbon nanotubes (VA-CNTs, both single-walled and multiwalled) show great potential for dry adhesive applications. Although an adhesive strength of more than 500 N cm^{-2} between VA-CNTs and a glass surface has been predicted by theory (7, 8), experimental work carried out so far showed rather low adhesion forces up to only about 30 N cm^{-2} for macroscopic arrays of vertically aligned single-walled carbon nanotubes (VA-SWNTs) (9) and 36 N cm^{-2} for micropatterned arrays of vertically aligned multiwalled carbon nanotubes (VA-MWNTs) (10). Recent atomic force microscopic (AFM) measurements have revealed a strong nanometer-scale adhesion with AFM tips up to 200 or 3 to 15 times stronger than that offered by gecko foot hairs for VA-MWNTs or functionalized polymer pillars (11, 12). The large difference between the observed macroscopic adhesion forces for VA-CNTs and the theoretical prediction is presumably due to their inaccessibility to the hierarchical structure of geckos' setae and spatulae. Moreover, a strong lift force is normally required to detach carbon nanotube dry adhesives that strongly bind to a surface, which limits the application of VA-CNTs as a transient adhesive.

Theoretical studies have indicated that an optimal adhesion could be achieved by the combination of a size reduction and shape optimization with hierarchical structures (13, 14) and that the side contact of fibers with substrates over a larger contact area could provide a stronger adhesion force than that of a tip contact (15, 16). With use of hierarchically structured VA-CNT arrays, we report here gecko-foot–mimetic dry adhesives with a high shear adhesion force (~100 N cm^{-2}) for strong shear binding-on but a much lower normal adhesion force (~10 N cm^{-2}) for easy normal lifting-off. The carbon nanotube arrays are required to have a straight aligning body and a curly entangled end segment at the top. This is responsible for creating an anisotropic adhesion force through the sidewall contact with various substrates, and the difference between shear adhesion and normal adhesion is what the gecko exploits to switch between attachment and detachment as it moves. The required samples were produced by a low-pressure chemical vapor deposition (CVD) process on a SiO$_2$/Si wafer (fig. S1) (9, 17–19). During the pyrolytic growth of the VA-MWNTs, the initially formed nanotube segments from the base growth process grew in random directions and formed a randomly entangled nanotube top layer to which the underlying straight nanotube arrays then emerged (7, 20) [supporting online material (SOM)]. However, Zhao et al. (7) suggested that the top layer of the entangled nanotube segments could prevent the underlying aligned nanotubes from contacting the target surface, leading to a weakened adhesion force.

To demonstrate the adhesion performance of the VA-MWNTs, we finger-pressed a small piece of the as-grown VA-MWNT film (4 mm by 4 mm, Fig. 1A) from the Si side onto a vertically positioned glass slide. The nanotubes in this film have diameters ranging from 10 to 15 nm with a tube length of about 150 μm and a tube density of ~10^{10} to 10^{11} cm^{-2} (Fig. 1, B and C). A book of 1480 g was clung onto a thin wire that was preglued on the back side of the SiO$_2$/Si substrate. An overall adhesion force of 90.7 N cm^{-2} was calculated for the VA-MWNT dry adhesive film shown in Fig. 1A. Similar adhesion behaviors were observed for the VA-MWNT dry adhesive against various other substrates with different flexibilities and surface characteristics, including ground glass plates, polytetrafluoroethylene (PTFE) film, rough sandpaper, and poly(ethylene terephthalate) (PET) sheet (figs. S3 to S6).

As shown in Fig. 1D, the normal adhesion force for VA-MWNT films with the tube length ranging
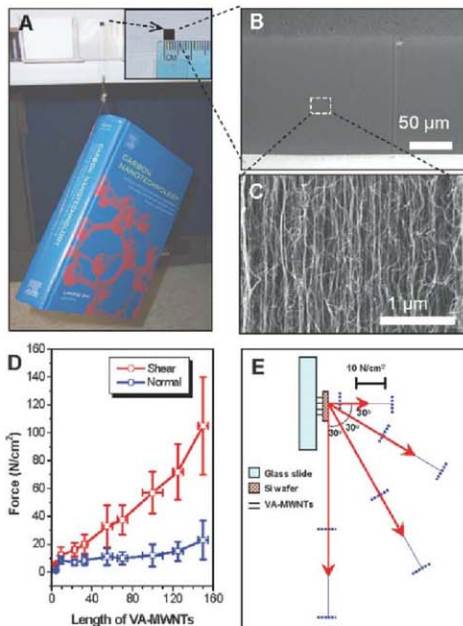
Fig. 1. (**A**) A book of 1480 g in weight suspended from a glass surface with use of VA-MWNTs supported on a silicon wafer. The top right squared area shows the VA-MWNT array film, 4 mm by 4 mm. (**B** and **C**) SEM images of the VA-MWNT film under different magnifications. (**D**) Nanotube length–dependent adhesion force of VA-MWNT films attached onto the substrate with a preloading of 2 kg (7, 9). The vertical and horizontal bars represent the deviations of the force and the nanotube length, respectively, measured for more than 20 samples of the same class. (**E**) Adhesion strength of VA-MWNTs with length 100 ± 10 μm at different pull-away directions. The red arrows represents the average forces measured for more than 20 samples, whereas the two perpendicular blue dot lines define possible deviations of the force measured for different samples of the same class. The nanotubes and substrates shown in (E) are not to scale.

[1]Department of Chemical and Materials Engineering, School of Engineering, University of Dayton, 300 College Park, Dayton, OH 45469, USA. [2]Air Force Research Laboratory, Human Effectiveness Directorate, AFRL/RH, Wright-Patterson Air Force Base, OH 45433, USA. [3]Department of Mechanical Engineering, University of Akron, Akron, OH 44325, USA. [4]School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA.

*To whom correspondence should be addressed. E-mail: ldai@udayton.edu (L.D.); zlwang@gatech.edu (Z.L.W.)

from about 10 to 150 μm increased slightly from 10 to 20 N cm⁻². However, the corresponding shear adhesion force increased from 10 to 100 N cm⁻² over the same range of nanotube lengths. The shear adhesion force is typically several times stronger than the corresponding normal adhesion force at a constant nanotube length over about 10 μm. The high shear adhesion force of the VA-MWNT dry adhesive ensures a strong adhesion to the target surface for hanging heavy objects along the shear direction, whereas a much weaker normal adhesion force allows the nanotube film to be readily detached in the normal direction. As shown in movie S1, a finger-tip press can firmly attach a VA-MWNT film (~100-μm tube length, 4 mm by 4 mm in area), supported by a SiO₂/Si substrate used for the nanotube growth, onto a vertically positioned glass slide to hold an ~1000-g weight [400-g beaker (Pyrex, Corning Incorporated, Corning, New York, 1000 ml) plus 600 ml water] in the shear direction. The VA-MWNT arrays were repeatedly attached and detached from the glass surface, and the supported weight did not decrease. The VA-MWNT dry adhesive was found to spontaneously peel away from the glass slide upon tilting it toward the horizontal level with the loaded object facing downward (movie S2). This observation is consistent with the easy detachment of a gecko foot at a tilted angle from a target surface (1, 14). To elucidate the angular dependence of the adhesion forces, we measured the pull-off force in various pull-away directions. The decrease in the pull-off force with increasing pull-away angle shown in Fig. 1E indicates that the shear adhesion force is much stronger than the normal adhesion force.

Because of the minimal hydrogen bonding (fig. S7) and negligible electrostatic charging effects (fig. S8), the vdW force is mainly responsible for the adhesive force between the nanotube film and the glass slide (2). As such, the structure at and near the top surface of the VA-MWNT film plays a critical role in regulating its adhesive performance. We examined the morphology of the top surface and cross-sectional area of the VA-MWNT films with different tube lengths before and after the shear adhesion measurements. As expected, randomly entangled nanotube segments arising from the initial stage of the base growth process were observed on the top surface of the as-synthesized VA-MWNT arrays (Fig. 2A, a to c). After the shear adhesion force measurements, however, we found that the top layer of the randomly entangled nanotube segments became horizontally aligned (Fig. 2A, d to f). The degree for the shear-induced horizontal alignment increased with increasing the aligned nanotube length (Fig. 2A, d to f). Before the testing, the nanotube "trunks" were uniformly aligned (Fig. 2A, g to i). However, after binding on the wall, the vertically aligned nanotube trunks were tilted along the shear direction (Fig. 2A, j to l). The significant increase in the shear adhesion force with the aligned nanotube length observed in Fig. 1D seems

to be directly related to the presence of the horizontally aligned nanotube segments on the top surface of the VA-MWNT dry adhesive films, which formed the tube-length–dependent horizontally aligned structure under shear.

To prove the importance of the nonaligned entanglement at the top of the VA-MWNTs, we carried out experiments with arrays prepared by a conventional CVD process without a vacuum system (19). The adhesion forces are generally less than 1 N/cm² because of the absence of a nonaligned nanotube top layer and/or its poor quality (fig. S18). Both the nanotube structural defects and amorphous carbon contaminants were found to significantly reduce the adhesion forces (figs. S16 and S17).

The scanning electron microscopy (SEM) observations are consistent with the following process. During the initial contact, the top non-
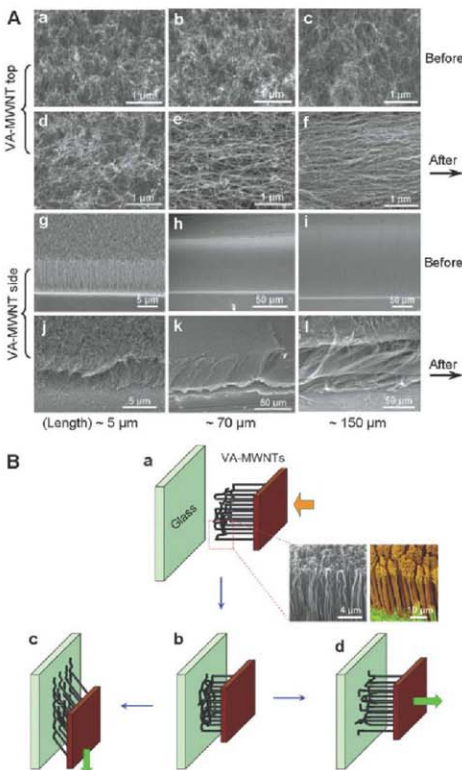
Fig. 2. SEM images and schematic diagrams for the morphological change of VA-MWNT arrays during adhesion measurements. (**A**) Top (a to f) and side (g to l) views of VA-MWNT films with different length before (a to c and g to i) and after (d to f and i to l) adhesion measurements. The VA-MWNT length in a, d, g, and j is ~5 μm; in b, e, h, and k, ~70 μm; and in c, f, i, and l, ~150 μm. The arrows indicate the shear direction during the shear adhesion force measurements. (**B**) Preloading (a), attachment of the VA-MWNT array onto the glass substrate (b), shear adhesion force stretching the nonaligned nanotubes on the substrate to form the line contact (c), and normal adhesion force leading to the nonaligned nanotubes point-by-point peel-off from the substrate (d). (Inset) The structure similarity between the cross-section views of the VA-MWNTs (left) and gecko's aligned elastic hairs (right).

aligned nanotube segments (Fig. 2Ba) adopted randomly distributed "line" contact with the glass substrate (Fig. 2Bb). Upon shear adhesion force measurement (Fig. 2Bc), the applied shear force caused the nonaligned nanotube segments to align along the shear direction on the glass substrate (Fig. 2Bc) and the vertically aligned nanotube trunks to tilt along the shear direction (Fig. 2A, j to l), leading to a predominant aligned

line contact with the glass surface (Fig. 2A, d to f). This is also consistent with the strong length dependence of the shear adhesion force shown in Fig. 1D, in which the longer VA-MWNTs offer support longer randomly entangled nanotube segments on the top for a more extensive line contact upon shear and offer more flexibility for titling the long nanotube trunks to achieve the most intimate contacts with the target substrate.

During the normal adhesion force measurements, however, the top nonaligned nanotube segments contacted with the glass substrate were peeled from the substrate through a "point-by-point" detaching process (Fig. 2Bd), requiring a much lower force than that for pulling off the entire nanotube array (Fig. 1D). These failure modes have been demonstrated by computer simulations (fig. S2 and movies S3 to S5). The line contact detachment (Fig. 2Bc) is expected to produce a stronger shear adhesion force than the normal adhesion force governed by the point-by-point peel-off detachment (Fig. 2Bd). The normal adhesion force only increased slightly (fig. S10) with the preshear–induced nanotube alignment, whereas the shear adhesion force increased much more dramatically (fig. S11). We also investigated the time dependence of the VA-MWNT dry adhesives compared with those of commercial copper adhesive tapes. As seen in fig. S9, the VA-MWNT dry adhesive (60 μm long) under a shear loading of 40 N cm^{-2} or a normal pull-away force of 12 N cm^{-2} for more than 24 hours remained on the glass substrate stably without any cohesive breakage. In contrast, commercial copper adhesive tapes (3M, St. Paul, Minnesota) under the same applied forces fatigued easily and were peeled away from the substrate within 1 hour.

We also used oxygen plasma etching (21, 22) to physically remove the nonaligned nanotube segments to investigate their influence on the adhesion forces (fig. S12, a to d). The removal of the nonaligned nanotube segments from the top of a VA-MWNT array by plasma etching led to predominately point contacts, which largely eliminated the nanotube-length–dependence for both shear and normal adhesion forces within the experimental error, with a concomitant decrease in adhesion forces (fig. S12e). The plasma-etching-induced "bundle" formation (fig. S12c) together with the associated surface chemistry changes (fig. S13) also weakened adhesion forces by reducing the number of effective contact points per unit area and/or the interaction energy per contact with the glass surface. To study the influence of nonaligned nanotube segments on the adhesion forces in a more precise way, we eliminated nonaligned nanotube segments from the top of a VA-MWNT array by turning over the as-synthesized VA-MWNT film with full integrity from the SiO$_2$/Si wafer onto a polystyrene substrate to keep the nanotube density and surface chemistry largely unchanged (figs. S14 and S15). The side view SEM image for the inverted VA-MWNT film similar to the sample shown in Fig. 2Ah is given in Fig. 3A under a relatively high magnification to show individual nanotubes without nonaligned top segment. After the adhesion force measurements, Fig. 3B shows the shear-induced alignment but with much fewer horizontally aligned nanotube segments on the top surface with respect to Fig. 2Ae. Figure 3C shows a significant decrease in the shear adhesion force with a slightly weakened normal adhesion force by inverting the VA-MWNT array. These results

Fig. 3. Typical side view of the inverted VA-MWNT film (fig. S14) without top entangled nanotube segments before (**A**) and after (**B**) adhesion measurements; (**C**) the shear and normal adhesions of VA-MWNT films with and without top entangled nanotube segments (nanotube length ~ 80 μm). Error bars represent the deviations of the forces measured for more than 20 samples of the same class.

Fig. 4. Experimental and predicted total adhesion force as a function of loading angle. The predictions show two failure modes: shear failure ($\theta < 25°$) and normal failure ($\theta > 25°$). (**Inset**) A schematic representation of a nanotube attached on a substrate.

indicate that the enhanced shear adhesion force for the VA-MWNT dry adhesive relies on the presence of curly entangled nanotube segments for large sidewall contact. Compared with VA-SWNT arrays with no nonaligned nanotube segments on the top [9], the VA-MWNT dry adhesive shows an expected much stronger shear adhesion force. However, the VA-SWNT dry adhesive possesses a normal adhesion force ($\sim$30 N cm^{-2}) higher than that of its VA-MWNT counterpart (<20 N cm^{-2}), which may be related to a higher nanotube graphitization degree and packing density associated with the VA-SWNT array [9].

The contributions of the top nonaligned carbon nanotubes to the adhesion can be estimated. Assuming that there is a line contact between the nonaligned nanotubes and the substrate with an effective contact length L after preloading, the VA-MWNT trunk forms an angle θ_0 with the substrate surface (Fig. 4 inset). We first consider the friction between the substrate and a single VA-MWNT. The attractive force per unit length on the VA-MWNT is $F_{vdw} = A\sqrt{d}/(16D^{2.5})$, with the Hamaker constant A, the nanotube diameter d, and the gap distance between the nanotube surface and the substrate D [23]. There is a cut-off gap distance $D = D_0$, representing the effective separation between the nanotube and the substrate, at which maximum attractive force, F_{vdw}^{max}, is estimated [23]. The maximum friction force per unit area of the VA-MWNT film is $P_f^{max} = \mu F_{vdw}^{max} L\rho$, where ρ is the effective nanotube contact density per unit area and μ is the friction coefficient. Taking the values of $L = 100$ nm, $\mu = 0.09$ [24], $\rho = 5 \times 10^{10}$ tubes cm^{-2}, $d = 15$ nm, $A = 6 \times 10^{-20}$ J [23], and $D_0 = 0.34$ nm [25], then P_f^{max} is 97 N cm^{-2}, a value close to the experimental results. L is only a fraction of the length of those observed nonaligned nanotube segments (over 1 μm) at the top of the VA-MWNT arrays, and hence more design space may exist for further increasing the adhesion force. The normal adhesion force of the VA-MWNT film can be calculated on the basis of the geometric relations shown in the Fig. 4 inset. The maximum attractive force per unit area can be obtained by integrating vdW force in the peel zone [26], $P_n^{max} = \rho\int_{x_0}^{x_1}A\sqrt{d}/(16D^{2.5})dx$, where point x_0 is the last contact point between the carbon nanotube and the substrate; point x_1 is the point beyond which the vdW force can be neglected, which occurs at a (critical) separation distance D_c ($D_c = 5D_0$ is used here). Assuming that the nanotube at the peel zone is curved with a radius of R_s (Fig. 4 inset), $D = D_0 + R_s - \sqrt{R_s^2 - x^2}$. If a load σ is applied on the nanotube array with an angle θ, the criteria for normal and shear failures are $\sigma > P_n^{max}/\sin\theta$ and $\sigma > P_f^{max}/\cos\theta$, respectively. Figure 4 shows the total forces as a function of θ, predicted by the above formulae, in which the unknown parameters R_s, ρ, and L are obtained by fitting the experimental data. There are normal and shear failure modes, depending on loading angle θ. At a small angle ($\theta < 25°$), the applied shear stress will first exceed

the interfacial shear strength P_f^{max}, resulting in shear failure, whereas at a large angle ($\theta > 25°$) the applied normal force will first exceed the normal strength P_n^{max}, leading to a detachment. These results are consistent with the finite element analysis for a gecko foot [27], indicating that the important anisotropic mechanism for gecko adhesives is mimicked by the VA-MWNT dry adhesives. It can be seen from the model that the parameters R_s, ρ, and L represent the geometric effect of the nonaligned nanotube segments on normal and friction forces, whereas the Hamaker constant A is the contribution of nanotube surface chemistry. Increasing the film thickness may increase the number of effective nanotube contacts (ρ) and length of line contact (L) and thus increase both normal and shear adhesion, as shown by the data in Fig. 1D. The clean surface of the VA-MWNT arrays produced in our study may lead to a relatively high Hamaker constant, which is consistent with the stronger adhesion forces observed for our VA-MWNTs compared with those for the nanotube arrays of a similar structure produced by a conventional CVD process reported in [7, 19] (figs. S16 and S18).

For VA-MWNTs without nonaligned top segments, the nanotubes may not have line contact with substrate. Assuming that all of the nanotubes contact with the substrate at their top ends, the attractive force per unit area on nanotubes is $F_{vdw} = \rho Ad/12D^2$ [23]. Using the above values of A, d, D, ρ, and $\mu = 0.8$ to 1.7 [26, 28], we calculated the maximum normal and shear forces per unit area to be 32 N cm^{-2} and 26 to 55 N cm^{-2}, respectively. However, it is likely that only a fraction of the nanotube tips contact with the substrate (lower ρ) because of bundle formation at the plasma-etched VA-MWNT tips, resulting in much lower values of normal and shear forces, as shown in fig. S12. For the inverted VA-MWNT array, the experiment data are close to the range of theoretical predictions (Fig. 3C) given that the nanotube tip may have a slightly different Hamaker constant A from that of its sidewall.

We designed VA-MWNT arrays to mimic gecko feet with a shear adhesive force of close to 100 N cm^{-2} while retaining a normal adhesion force comparable to that of gecko feet (about 10 N cm^{-2}). Shear-induced alignment of the nonaligned nanotube top layer dramatically enhanced the shear adhesion force resulting from line contact, which increased rapidly with increasing tube length. In contrast, the normal adhesion force is almost insensitive to the nanotube length as a result of point contact. An alternative sticking and detaching of the VA-MWNT on various substrates with different flexibilities and surface characteristics, including glass plates, PTFE film, rough sandpaper, and PET sheet, can mimic the walking of a living gecko. This finding enables us to construct aligned carbon nanotube dry adhesives with a strong shear adhesion for firm attachment and relatively weak normal adhesion for easy detachment, which opens many technological applications.

References and Notes

1. K. Autumn et al., Nature 405, 681 (2000).
2. K. Autumn et al., Proc. Natl. Acad. Sci. U.S.A. 99, 12252 (2002).
3. D. J. Irschick et al., Biol. J. Lin. Soc. 59, 21 (1996).
4. M. Sitti, R. S. Fearing, J. Adhes. Sci. Technol. 17, 1055 (2003).
5. A. K. Geim et al., Nat. Mater. 2, 461 (2003).
6. L. Dai, Ed., Carbon Nanotechnology, Recent Developments in Chemistry, Physics, Materials Science and Device Applications (Elsevier, Amsterdam, 2006).
7. Y. Zhao et al., J. Vac. Sci. Technol. B 24, 331 (2006).
8. M. F. Yu, T. Kowalewski, R. S. Ruoff, Phys. Rev. Lett. 86, 87 (2001).
9. L. Qu, L. Dai, Adv. Mater. 19, 3844 (2007).
10. L. Ge, S. Sethi, L. Ci, P. M. Ajayan, A. Dhinojwala, Proc. Natl. Acad. Sci. U.S.A. 104, 10792 (2007).
11. B. Yurdumakan, N. R. Raravikar, P. M. Ajayan, A. Dhinojwala, Chem. Commun. (Camb.) 30, 3799 (2005).
12. H. Lee, B. P. Lee, P. B. Messersmith, Nature 448, 338 (2007).
13. H. J. Gao, H. M. Yao, Proc. Natl. Acad. Sci. U.S.A. 101, 7851 (2004).
14. H. Yao, H. J. Gao, J. Mech. Phys. Solids 54, 1120 (2006).
15. J. Lee, B. Schubert, C. Majidi, R. S. Fearing, J. R. Soc. Interface 5, 835 (2008).
16. C. S. Majidi, R. E. Groff, R. S. Fearing, J. Appl. Phys. 98, 103521 (2005).
17. X. B. Zhang et al., Adv. Mater. 18, 1505 (2006).
18. L. T. Qu, L. M. Dai, J. Mater. Chem. 17, 3401 (2007).
19. It has been previously reported that a long MWNT array (e.g., 100 μm) showed an adhesion force lower than that of a short nanotube array (e.g., 5 to 10 μm) [7]. However, an independent study revealed that a long MWNT array (>100 μm) in a patterned fashion supported a stronger adhesion force than its short counterpart [10]. These results may indicate that the adhesion force of aligned nanotube arrays depends strongly on the nature of the nanotube sample and hence the nanotube growth conditions. To maximize the vdW interaction between the carbon nanotubes and substrate surfaces for strong adhesion, we must control the nanotube synthesis to ensure a clean surface for the resultant aligned nanotube array. Similar requirements have been demonstrated previously for the carbon nanotube yarn formation [17]. Therefore, we have exploited the low-pressure CVD process for the growth of VA-MWNT arrays free from amorphous carbon or other possible surface contamination. Furthermore, the base growth process led to the formation of hierarchically structured carbon nanotube dry adhesives consisting of randomly entangled nanotube segments on top of the resultant VA-MWNT array [7, 20].
20. S. M. Huang, L. Dai, A. W. H. Mau, J. Phys. Chem. B 103, 4223 (1999).
21. X. K. Lu, H. Huang, N. Nemchuk, R. S. Ruoff, Appl. Phys. Lett. 75, 193 (1999).
22. S. M. Huang, L. Dai, J. Phys. Chem. B 106, 3543 (2002).
23. D. Leckband, J. Israelachvili, Q. Rev. Biophys. 34, 105 (2001).
24. P. L. Dickrell et al., Tribol. Lett. 18, 59 (2005).
25. S. Akita, Y. Nakayama, Jpn. J. Appl. Phys. 41, 4242 (2002).
26. Y. Tian et al., Proc. Natl. Acad. Sci. U.S.A. 103, 19320 (2006).
27. H. Gao, X. Wang, H. Yao, S. Gorb, E. Arzt, Mech. Mater. 37, 275 (2005).
28. H. Kinoshita, I. Kume, M. Tagawa, N. Ohmae, Appl. Phys. Lett. 85, 2780 (2004).
29. L.D., Z.L.W., and M.S. thank AFRL/Air Force Office of Scientific Research for financial support. L.D. also thanks T. Yamada, S. Sangwook, A. Roy, J. Baur, and T. Benson-Tolle for useful discussions as well as financial support from NSF (grant CMS-0609077).

Base Sequence and Higher-Order Structure Induce the Complex Excited-State Dynamics in DNA

Nina K. Schwalb and Friedrich Temps

The high photostability of DNA is commonly attributed to efficient radiationless electronic relaxation processes. We used femtosecond time-resolved fluorescence spectroscopy to reveal that the ensuing dynamics are strongly dependent on base sequence and are also affected by higher-order structure. Excited electronic state lifetimes in dG-doped d(A)$_{20}$ single-stranded DNA and dG·dC-doped d(A)$_{20}$·d(T)$_{20}$ double-stranded DNA decrease sharply with the substitution of only a few bases. In duplexes containing d(AGA)·d(TCT) or d(AG)·d(TC) repeats, deactivation of the fluorescing states occurs on the subpicosecond time scale, but the excited-state lifetimes increase again in extended d(G) runs. The results point at more complex and molecule-specific photodynamics in native DNA than may be evident in simpler model systems.

The high photostability of DNA is of supreme importance for the stability of genetic information. It is commonly attributed to efficient radiationless deactivation processes, whereby the energy of an absorbed ultraviolet (UV) photon is rapidly dissipated before chemical reactions in the excited state can cause profound photochemical damage (1). The underlying photophysical mechanisms in DNA are, however, only poorly understood. In particular, it is unknown why excited electronic-state lifetimes in DNA can be some surprising three to four orders of magnitude longer (2–5) than the ultrashort (subpicosecond) lifetimes of the free nucleobases adenine (A), thymine (T), guanine (G), and cytosine (C) (6–9) that are the building blocks of DNA.

The complexity of the matter is underscored by recent findings pointing at very different roles of hydrogen (H) bonding (10–13) versus base-stacking interactions in DNA (2, 3, 14–16). Electronic relaxation in isolated H-bonded guanosine-cytidine (G·C) Watson-Crick pairs is strongly accelerated (11) by a G-to-C charge transfer followed by rapid intermolecular proton migration and subsequent hydrogen back-transfer to the ground state (12, 13). In contrast, measurements of d(A)$_n$ single-stranded DNA (ssDNA) and d(A)$_n$·d(T)$_n$ or d(AT)$_{n/2}$·d(TA)$_{n/2}$ double-stranded DNA (dsDNA) oligonucleotides with $n \geq 20$ bases or polymers ($n \approx 200$ to 2000) showed excited-state lifetimes reaching into the 100-ps and even 1-ns range (2–5).

The explanation for this very puzzling result is a matter of substantial controversy (17, 18). It is generally suspected that the prime cause is the existence of delocalized excited states, but it is questionable whether the observed dynamics arise from transformations of the initially excited states to exciplex states with partial charge-transfer char-

acter (3, 19) or whether exciton states with delocalized oscillator strengths over a small number of stacked bases are involved, and the dynamics just reflect the wide spectrum of the resulting eigenstates (5, 14–16, 20, 21). Moreover, it is important to know to what extent the excitation remains within one strand or is distributed over both strands. Base flip-out (3) and contributions of static and dynamic disorder (22, 23) are intensely debated as well. In the broader context, the simple d(A)$_n$ and d(AT)$_{n/2}$ strands studied to date have very atypical sequences, and it is unclear how strictly those data apply to native DNA. In order to address these questions, it is necessary to investigate the sequence dependence of the complex excited-state dynamics in DNA as well as possible influences by higher-order (secondary, tertiary, and quaternary) structural variations.

We studied a series of dC-doped d(T)$_{20}$ and dG-doped d(A)$_{20}$ DNA oligonucleotides and their resulting duplexes by means of femtosecond fluorescence spectroscopy. The experiments probed the relaxation dynamics of the optically bright excited states of the molecules, which may return to their electronic ground states in principle either directly or via optically dark intermediate states that deactivate to the ground states in longer times. The sequences are displayed in Table 1. The samples were supplied in several small batches of ≈0.8 μmol each and were dissolved in an aqueous phosphate buffer, thermally annealed, and adjusted to concentrations of 0.46 mM per base (0.023 mM of the 20-nucleotide oligomers) using standard protocols (24). Duplexes were prepared by mixing equimolar amounts of the single strands. All solutions were characterized by UV absorption spectra, melting curves, and circular dichroism (CD) spectra (24). Time-resolved fluorescence measurements were carried out using the upconversion technique (11) at a fluorescence wavelength (λ_{fl}) of 350 nm in a 1-mm flow cell at room temperature. UV excitation pulses of ≤0.1 μJ/pulse and ≈50 fs full width at half maximum duration at an excitation wavelength (λ_{pump}) of 269 nm were

provided by a frequency-doubled noncollinear optical parametric amplifier. Great care was taken to rule out photodamage of the sensitive samples by using low pump powers and short measurement times (<0.1 mW average power and <40 min per decay curve). The time resolution was about 200 fs.

We first investigated the pyrimidine-base ssDNA molecules d(T)$_{20}$, d(TTCTT)$_4$, dT(TCT)$_6$T, d(TC)$_{10}$, d(TC)$_9$T, d(C)$_{20}$, and d(C)$_{10}$. The measured fluorescence time profiles (Fig. 1, A and B) for these strands exhibit quite similar biexponential behaviors with subpicosecond lifetimes (τ_1) describing the main initial decays and weaker secondary components (τ_2) in the few-picosecond range. A fit to the transients (25) showed time constants of $\tau_1 = 0.60 \pm 0.04$ (2 SEM) ps and $\tau_2 = 2.96 \pm 0.40$ ps for the cases of d(T)$_{20}$ and d(TTCTT)$_4$ and $\tau_1 = 0.33 \pm 0.01$ ps and $\tau_2 = 1.74 \pm 0.11$ ps for the cases of dT(TCT)$_6$T, d(TC)$_{10}$, d(TC)$_9$T, and d(C)$_{20}$. The decay curve for d(C)$_{20}$ appears to be slightly faster than the others because of a smaller amplitude of its second component.

Next in line were the pure and mixed purine-base single strands d(A)$_{20}$, d(AAGAA)$_4$, dA(AGA)$_6$A, d(AG)$_{10}$, d(AG)$_9$A, d(G)$_{20}$, and d(G)$_{10}$. In contrast to the pyrimidine strands, their fluorescence-decay profiles vary profoundly with the sequences (Fig. 1C). Furthermore, they require three exponentials for satisfactory fitting and share a long-lived fluorescence component with a time constant $\tau_3 = 97 \pm 20$ ps determined by a global fit. Its amplitude is healthy (7%) in d(A)$_{20}$ and d(G)$_{20}$ but drops to ≈1 to 2% in the mixed decamer d(G)$_{10}$ and even falls out of the fit within our experimental error limits for the alternating d(AG)$_9$A. Figure 1D displays the respective time profiles on an extended time scale to show the long-lived contributions. The fast subpicosecond (τ_1) and intermediate several-picosecond (τ_2) decay components vary. Virtually identical values of $\tau_1 = 0.63 \pm 0.03$ ps and $\tau_2 = 5.8 \pm 0.4$ ps are found for d(A)$_{20}$ and d(AAGAA)$_4$, although the overall decay of the excited d(AAGAA)$_4$ is much faster because it has almost lost its τ_3 component. Continuing to dA(AGA)$_6$A, the fluorescence decay rate reaches a pronounced maximum, with $\tau_1 = 0.30 \pm 0.04$ ps and $\tau_2 = 2.3 \pm 0.4$ ps. In contrast, as the dG content is raised further, in the alternating sequences d(AG)$_{10}$ and d(AG)$_9$A, the lifetimes increase again to the former values of $\tau_1 = 0.63 \pm 0.03$ ps and $\tau_2 = 5.8 \pm 0.4$ ps. Recorded CD spectra (fig. S2) indicate that this effect coincides with emerging structural changes caused by substantial G stacking and quadruplex formation. The increase is even more substantial in pure d(G)$_{20}$ [as well as d(G)$_{10}$], where $\tau_1 = 1.17 \pm 0.12$ ps and $\tau_2 = 10.4 \pm 1.1$ ps [0.83 ± 0.06 and 10.4 ± 0.11 ps in d(G)$_{10}$], nearly twice as long as for d(AG)$_{10}$. The amplitudes of the τ_1 and τ_2 components also vary substantially.

The temporal fluorescence profiles of the complementary duplexes are displayed in Fig. 1E. In addition to a subpicosecond decay component (τ_1), all double strands share a component in

Institut für Physikalische Chemie, Christian-Albrechts-Universität zu Kiel, Olshausenstrasse 40, D-24098 Kiel, Germany. E-mail: schwalb@phc.uni-kiel.de (N.K.S.); temps@phc.uni-kiel.de (F.T.)

the picosecond range [$\tau_2 = 2.6 \pm 0.3$ ps, except d(A)$_{20}$, for which $\tau_2 = 1.6 \pm 0.4$ ps] and a longer-lived decay component ($\tau_3 = 16 \pm 4$ ps).

As is illustrated by Fig. 1F, the latter is most notable in d(G)$_{20}$·d(C)$_{20}$ (15%), d(G)$_{10}$·d(C)$_{10}$ (8%), and d(A)$_{20}$·d(T)$_{20}$ (7%) but absent or

nearly absent in the mixed duplexes (0 to 2%). Initial subpicosecond decay processes with $\tau_1 = 0.52 \pm 0.04$ ps (40 to 70% contributions) take place in d(A)$_{20}$·d(T)$_{20}$ and in the A-T-rich duplex d(AAGAA)$_4$·d(TTCTT)$_4$ but also in d(G)$_{20}$·d(C)$_{20}$ and d(G)$_{10}$·d(C)$_{10}$. In contrast to d(A)$_{20}$·d(T)$_{20}$ and d(G)$_{10,20}$·d(C)$_{10,20}$, however, the more highly doped dA(AGA)$_6$A-dT(TCT)$_6$T and especially the alternating sequences d(AG)$_{10}$·d(TC)$_{10}$ and d(AG)$_9$A·d(TC)$_9$T stand out for their ultrafast initial electronic relaxation rates. They almost lack the long-lived τ_3 component but share a very short first-time constant of $\tau_1 = 0.34 \pm 0.02$ ps. Carrying ≈80 to 90% of the amplitude, ultrafast processes clearly dominate the fluorescence dynamics in those strands.

The mean lifetimes of the fluorescing excited states of the different DNA strands are compared in Table 2. As can be seen, the results for the pure d(T)$_{20}$ and mixed d(T,C)$_{20}$ pyrimidine single strands differ little; only the values for d(C)$_{10,20}$ look substantially shorter. The CD spectra (fig. S2) show normal B-form DNA (B-DNA) shape, with some A-form DNA (A-DNA)–like features only in d(C)$_{10,20}$. All lifetimes are slightly longer but still similar to those of the monomeric nucleotides (26). Extending a previous report for pure d(T)$_{18}$ (3), this finding leads to the conclusion that pyrimidine-base single strands do not generally form long-lived exciplex states.

Except for d(A)$_{20}$, for which our results are comparable with literature reports (2–4), mixed d(A,G) purine-base single strands have not been studied before in this context. From Table 2, their mean fluorescence lifetimes are strongly sequence-dependent. The data demonstrate unambiguously that the mean lifetimes of the fluorescing excited states in dG-doped d(A)$_n$ ssDNA drop sharply with the substitution of only every fifth dA by dG. Thus, at least the early electronic relaxation dynamics of the optically bright states that are probed by the transient fluorescence measurements are strongly accelerated in these strands. This important result was confirmed by measurements at $\lambda_{pump} = 284$ nm, closer to the electronic origins of the excited states (fig. S4). The CD spectra (fig. S2) for d(A)$_{20}$, d(AAGAA)$_4$, and dA(AGA)$_6$A show pronounced so-called exciton bands from base stacking in a normal B-DNA architecture (27), leading to the assumption that their initial optically excited states have substantial exciton character, as has been supposed for the case of pure d(A)$_n$ (5, 16). Because the oscillator strength in the dA-rich d(AAGAA) should still be governed by the dAs, there must be a sufficiently strong coupling between the bases in these stacks to explain the observed accelerated relaxation. Fast energy transfer between exciton states has indeed been proposed to take place in less than 100 fs (5, 21).

In view of the fact that G has the lowest redox potential of the natural nucleobases, it appears plausible that the faster excited-state relaxation in d(AAGAA)$_4$ and dA(AGA)$_6$A is due to a rapid

Table 1. Sequences of the investigated DNA oligonucleotides.

Base sequence	Abbreviation
5′-d(AAAAAAAAAAAAAAAAAAAA)-3′	d(A)$_{20}$
3′-d(TTTTTTTTTTTTTTTTTTTT)-5′	d(T)$_{20}$
5′-d(AAGAAAAGAAAAGAAAAGAA)-3′	d(AAGAA)$_4$
3′-d(TTCTTTTCTTTTCTTTTCTT)-5′	d(TTCTT)$_4$
5′-d(AAGAAGAAGAAGAAGAAGAA)-3′	dA(AGA)$_6$A
3′-d(TTCTTCTTCTTCTTCTTCTT)-5′	dT(TCT)$_6$T
5′-d(AGAGAGAGAGAGAGAGAGAG)-3′	d(AG)$_{10}$
3′-d(TCTCTCTCTCTCTCTCTCTCT)-5′	d(CT)$_{10}$
5′-d(AGAGAGAGAGAGAGAGAGA)-3′	d(AG)$_9$T
3′-d(TCTCTCTCTCTCTCTCTCT)-5′	d(TC)$_9$T
5′-d(GGGGGGGGGGGGGGGGGGGG)-3′	d(G)$_{20}$
3′-d(CCCCCCCCCCCCCCCCCCCC)-5′	d(C)$_{20}$
5′-d(GGGGGGGGGG)-3′	d(G)$_{10}$
3′-d(CCCCCCCCCC)-5′	d(C)$_{10}$
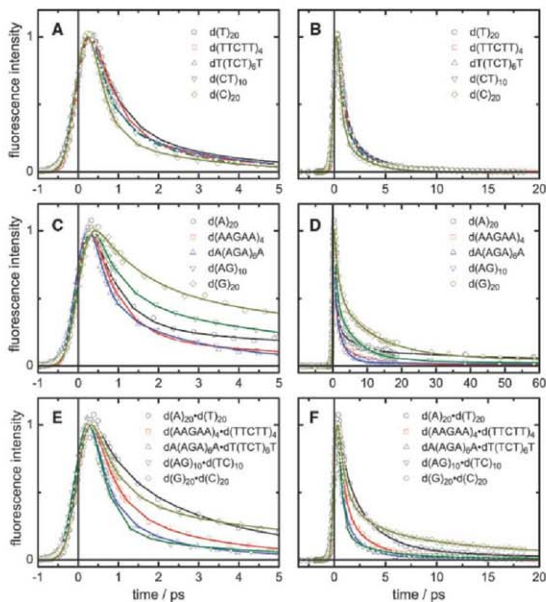
Fig. 1. Measured fluorescence-decay curves of the investigated DNA oligonucleotides on different time scales. (**A** and **B**) Pyrimidine-base single strands, (**C** and **D**) purine-base single strands, and (**E** and **F**) double strands. The time profiles for the alternating 19-nucleotide oligomers and 10-nucleotide oligomers are very similar to those of the respective 20-nucleotide oligomers and can be found in fig. S3.

Table 2. Mean fluorescence lifetimes* of the investigated ssDNA and dsDNA molecules.

Pyrimidine single strands		Purine single strands		Complementary double strands	
Sequence	$\langle \tau \rangle$ / ps	Sequence	$\langle \tau \rangle$ / ps	Sequence	$\langle \tau \rangle$ (ps)
$d(T)_{20}$	1.14	$d(A)_{20}$	8.05	$d(A)_{20} \cdot d(T)_{20}$	2.73
$d(TCTT)_4$	1.03	$d(AAGAA)_4$	3.54	$d(AAGAA)_4 \cdot d(TTCTT)_4$	1.35
$d(TCT)_6T$	0.74	$dA(AGA)_6A$	1.75	$dA(AGA)_6A \cdot dT(TCT)_6T$	0.76
$d(TC)_{10}$	0.90	$d(AG)_{10}$	3.78	$d(AG)_{10} \cdot d(TC)_{10}$	0.84
$d(TC)_9T$	0.90	$d(AG)_9A$	2.34	$d(AG)_9A \cdot d(TC)_9T$	0.63
$d(C)_{20}$	0.60	$d(G)_{20}$	11.3	$d(G)_{20} \cdot d(C)_{20}$	3.29
$d(C)_{10}$	0.53	$d(G)_{10}$	4.87	$d(G)_{10} \cdot d(C)_{10}$	2.07

*Average values of the fluorescence-decay components $\tau_{1,2,3}$ weighted by amplitude.

trapping of the hole created by the UV excitation at a G site and thereby to charge-transfer within the stacks. This point is of substantial interest also in the context of electron-hole transport in DNA (28–30). Our measurements show weakly fluorescing long-lived states in the purine single strands, and according to transient absorption measurements (2, 3), the long ≈100-ps electronic lifetimes are the signature of exciplex states with (partial) charge-transfer character within the π-stacks that are populated during the relaxation of the optically bright initially excited states. The optically dark character of the exciplex states is consistent with the very low emission amplitudes in our measurements. On the other hand, the longer excited state lifetimes in $d(AG)_{10}$ or $d(AG)_9A$ and in $d(G)_{10,20}$ ssDNA clearly come along with distinctive structural changes. According to their CD spectra (fig. S2), extended $d(AG)$ repeats show substantial G stacking and evidence for G quadruplex structures, whereas $d(G)_{10}$ and $d(G)_{20}$ exhibit even more profound interactions between their bases as well as emerging A-DNA–like features. Shorter distances between stacked bases or modified H-bonding motifs in quadruplex structures appear to affect the excited-state dynamics in subtle ways, which need to be elucidated by future work.

Duplex formation by Watson-Crick pairing overruns the looser higher-order structure of the single strands. CD spectra of the duplexes (fig. S2) exhibit normal B-DNA structures with strong exciton characteristics in the case of $d(A)_{20} \cdot d(T)_{20}$ but progressively weaker exciton features in $d(AAGAA)_4 \cdot d(TTCTT)_4$ and $dA(AGA)_6A \cdot dT(TCT)_6T$. The alternating $d(AG)_{10} \cdot d(TC)_{10}$ and $d(AG)_9A \cdot d(TC)_9T$ duplexes form B-DNA as well but lack the exciton band in their CD spectra, whereas $d(G)_{10,20} \cdot d(C)_{10,20}$ stand out for pronounced G stacking/quadruplex and some A-DNA–like features. The duplexes with extended $d(AGA)_n \cdot d(TCT)_n$ and $d(AG)_n \cdot d(TC)_n$ runs exhibit subpicosecond mean lifetimes and ultrafast initial decay components ($\tau_1 = 0.34$ ps) carrying >75% of the amplitudes. Thus, their excited-state relaxation dynamics as observable by fluorescence resembles the dynamics in the free mononucleotides. That unexpected outcome has great bearing for our understanding of the photophysics and photochemistry of native DNA. Supporting measurements at $\lambda_{pump} = 284$

nm showed practically identical behavior, so that the dynamics appear to be independent of excitation wavelength (fig. S4).

Pertaining to the ensuing ultrafast relaxation, the coupled electron-proton transfer mechanism established for the G·C Watson-Crick base pair (11–13) comes to mind as a very efficient electronic deactivation pathway in addition to the known subpicosecond relaxation pathways in the free bases. It is of interest in this context that a recent fluorescence measurement for poly-d(GC)·poly-d(GC) showed faster decay than in the mononucleotides (31). On the other hand, exciton or exciplex contributions are assumed to cause the long (≈16 ps) excited state lifetimes in extended $d(A)_n \cdot d(T)_n$ and, even more strongly, in $d(G)_n \cdot d(C)_n$ runs, although those lifetimes are shorter than in the single strands. A recent transient absorption study of long-lived ($\tau = 20$ to 40 ps) excited states in $d(G) \cdot d(C)$ DNA duplexes and hairpins (32) points at stepwise electronic deactivation of the initially excited optically bright states to the ground state via optically dark, nonfluorescent exciplex states. Although a coupled electron-proton motion such as in isolated G·C Watson-Crick pairs (11–13) was not assumed in those cases, extended $d(G)_n \cdot d(C)_n$ runs differ in structure from our more dA·dT-rich sequences, and a G·C-specific deactivation pathway in duplexes like $dA(AGA)_6A \cdot dT(TCT)_6T$ remains possible.

We have shown that the electronic deactivation dynamics in DNA are strongly dependent on base sequence and are affected as well by secondary, tertiary, and even quaternary structure, especially at high-dG content. The lifetimes of the fluorescing excited states in dG-doped $d(A)_{20}$ ssDNA and dG·dC-doped $d(A)_{20} \cdot d(T)_{20}$ dsDNA drop strikingly with the substitution of only a few sites. A limit is reached when couplings between adjacent guanines become so strong that G stacks and quadruplex formation come into play. Despite the differences, dynamical processes appear to take place in the DNA strands on several common characteristic time scales. Lifetimes of fluorescing excited states of tens or even hundreds of picoseconds are concluded to be rather atypical in native DNA. Further work is necessary to elucidate relaxation pathways that lead to long-lived optically dark intermediate states.

References and Notes
1. C. E. Crespo-Hernández, B. Cohen, P. M. Hare, B. Kohler, Chem. Rev. 104, 1977 (2004).
2. C. E. Crespo-Hernández, B. Kohler, J. Phys. Chem. B 108, 11182 (2004).
3. C. E. Crespo-Hernández, B. Cohen, B. Kohler, Nature 436, 1141 (2005).
4. D. Markovitsi, A. Sharonov, D. Onidas, T. Gustavsson, Chem. Phys. Chem. 4, 303 (2003).
5. D. Markovitsi, D. Onidas, T. Gustavsson, F. Talbot, E. Lazzarotto, J. Am. Chem. Soc. 127, 17130 (2005).
6. J.-M. Pecourt, J. Peon, B. Kohler, J. Am. Chem. Soc. 123, 10370 (2001).
7. J. Peon, A. H. Zewail, Chem. Phys. Lett. 348, 255 (2001).
8. T. Gustavsson, A. Sharonov, D. Markovitsi, Chem. Phys. Lett. 351, 195 (2002).
9. T. Pancur, N. K. Schwalb, F. Renth, F. Temps, Chem. Phys. 313, 199 (2005).
10. A. Abo-Riziq et al., Proc. Natl. Acad. Sci. U.S.A. 102, 20 (2005).
11. N. K. Schwalb, F. Temps, J. Am. Chem. Soc. 129, 9272 (2007).
12. A. L. Sobolewski, W. Domcke, C. Hättig, Proc. Natl. Acad. Sci. U.S.A. 102, 17903 (2005).
13. G. Groenhof et al., J. Am. Chem. Soc. 129, 6812 (2007).
14. B. Bouvier, T. Gustavsson, D. Markovitsi, P. Millié, Chem. Phys. 275, 75 (2002).
15. B. Bouvier et al., J. Phys. Chem. B 107, 13512 (2003).
16. E. Emanuele, D. Markovitsi, P. Millié, K. Zakrewska, Chem. Phys. Chem. 6, 1387 (2005).
17. D. Markovitsi et al., Nature 441, E7 (2006).
18. C. E. Crespo-Hernández, B. Cohen, B. Kohler, Nature 441, E8 (2006).
19. W.-M. Kwok, C. Ma, D. L. Phillips, J. Am. Chem. Soc. 128, 11894 (2006).
20. A. Czader, E. R. Bittner, J. Chem. Phys. 128, 035101 (2008).
21. K. Hyeon-Deuk, Y. Tanimura, M. Cho, J. Chem. Phys. 128, 135102 (2008).
22. D. Onidas, T. Gustavsson, E. Lazzarotto, D. Markovitsi, J. Phys. Chem. B 111, 9644 (2007).
23. I. Buchvarov, Q. Wang, M. Raytchev, A. Trifonov, T. Fiebig, Proc. Natl. Acad. Sci. U.S.A. 104, 4794 (2007).
24. Materials and methods are available as supporting online material on Science Online.
25. All measured fluorescence-decay curves were first individually fitted by sums of up to three exponentials. However, because the individual fit results showed nearly identical decay constants for $d(T)_{20}$ and $d(TTCTT)_4$ and for $d(TCT)_6T$, $d(TC)_{10}$, $d(TC)_9T$, and $d(C)_{20}$ within the standard deviations, we decided to perform a joint fit to the respective transients with the rationale being to determine more-precise values for those decay times and to elucidate possible similarities of the dynamics. Common time constants were also identified, when the time profiles for the purine single strands and the double strands were individually fitted, and joint fits to the respective data were therefore performed for those strands as well. The fitted time profiles and time constants and amplitudes are given in fig. S3 and table S1.
26. D. Onidas, D. Markovitsi, S. Marguet, T. Gustavsson, J. Phys. Chem. B 106, 11367 (2002).
27. V. A. Bloomfield, D. M. Crothers, I. Tinoco, Nucleic Acids: Structures, Properties and Functions (University Science Books, Sausalito, CA, 2000).
28. S. O. Kelley, J. K. Barton, Science 283, 375 (1999).
29. C. Wan, T. Fiebig, O. Schiemann, J. K. Barton, A. H. Zewail, Proc. Natl. Acad. Sci. U.S.A. 97, 14052 (2000).
30. B. Elias, F. Shao, J. K. Barton, J. Am. Chem. Soc. 130, 1152 (2008).
31. F.-A. Miannay, A. Bányász, T. Gustavsson, D. Markovitsi, J. Am. Chem. Soc. 129, 14574 (2007).
32. C. E. Crespo-Hernández, K. de La Harpe, B. Kohler, J. Am. Chem. Soc. 130, 10844 (2008).
33. This work has been supported by the Deutsche Forschungsgemeinschaft. Furthermore, the authors thank J. Grötzinger for letting us take the CD spectra.

Implications of Magma Transfer Between Multiple Reservoirs on Eruption Cycling

Derek Elsworth,[1]* Glen Mattioli,[2] Joshua Taron,[1] Barry Voight,[1] Richard Herd[3]

Volcanic eruptions are episodic despite being supplied by melt at a nearly constant rate. We used histories of magma efflux and surface deformation to geodetically image magma transfer within the deep crustal plumbing of the Soufrière Hills volcano on Montserrat, West Indies. For three cycles of effusion followed by discrete pauses, supply of the system from the deep crust and mantle was continuous. During periods of reinitiated high surface efflux, magma rose quickly and synchronously from a deflating mid-crustal reservoir (at about 12 kilometers) augmented from depth. During repose, the lower reservoir refilled from the deep supply, with only minor discharge transiting the upper chamber to surface. These observations are consistent with a model involving the continuous supply of magma from the deep crust and mantle into a voluminous and compliant mid-crustal reservoir, episodically valved below a shallow reservoir (at about 6 kilometers).

Continuous and highly resolved geodetic and efflux records are available for only a few volcanoes. One of those is the Soufrière Hills volcano (SHV) on Montserrat, West Indies (1–3), which has been erupting since 1995. These data provide a window into deep processes contributing to stratovolcano behavior. We constrained magma migration with wide-aperture geodetic data supplemented by a well-documented extrusion record, and used these to explore the effect of deeply sourced fluxes on short-term eruption periodicity. The global positioning system (GPS) array is capable of capturing magmatic exchange to a depth comparable to the distance across geodetic stations (~11 km).

The SHV has followed a pattern of seismic crises separated by about 30 years (4). The most recent volcanoseismic crises, in the 1890s, 1930s, and 1960s, are interpreted as aborted eruptions, and the seismic crisis in the 1990s developed into the ongoing eruption. Phreatic activity began in July 1995 after several years of seismic unrest. The most recent eruption comprises a series of 2- to 3-year eruptive episodes and interspersed pauses lasting 1.5 to 2 years (1). An andesite dome grew continuously, in episode 1 from November 1995 until ~10 March 1998, followed by a pause with passive dome collapse ending in November 1999 (5). This cycle of growth of an active lava dome followed by a pause was repeated between December 1999 and mid-July 2003, followed by a pause lasting until October 2005 (6). Episode 3 began in October 2005 and ended in March 2007. A pause followed, which appears to have ended in August 2008 with continuing slow extrusion of lava on the western flank of the dome.

The inversion of ~1995–1997 GPS data suggests that the early magmas reside in a chamber at a depth of about 5 km (2, 7–9). Crystal phases in erupted magmas also imply that they

[1]College of Earth and Mineral Sciences, Penn State University, University Park, PA 16802, USA; and Institute of Advanced Studies, University of Western Australia, Perth, Australia. [2]Department of Geosciences, University of Arkansas, Fayetteville, AR 72701, USA. [3]School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK.

*To whom correspondence should be addressed. E-mail: elsworth@psu.edu
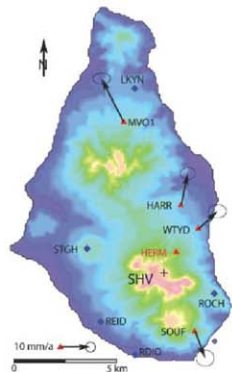
Fig. 1. Map of Montserrat, showing the location of the eruptive vent (black cross labeled SHV), continuous GPS sites (red triangles), and campaign GPS sites (blue diamonds) used in the flux analysis, and the Caribbean-fixed GPS velocity vectors for the period from 13 July 2003 through 1 November 2005, along with their 1σ errors for the continuous GPS (cGPS) sites. There is a strong radial deformation pattern relative to the vent, corresponding to inflation during this residual period (no surface magma flux). The proximal reference cGPS site HERM is shown in red.
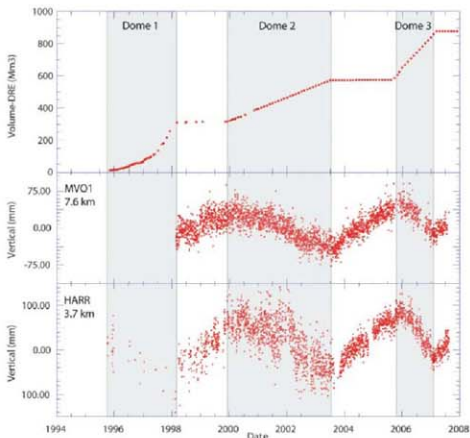
Fig. 2. Efflux of dense rock equivalent (DRE) from the SHV over time. Eruptive activity indicates three distinct active/repose cycles. Also shown is the evolution of station velocities within three prescribed cycles of activity. Resulting mean velocities are reported in table S1. Flux data from 1995 through early 1998 are from Sparks et al. (15) and data from 1998 are from electronically published MVO reports (6).
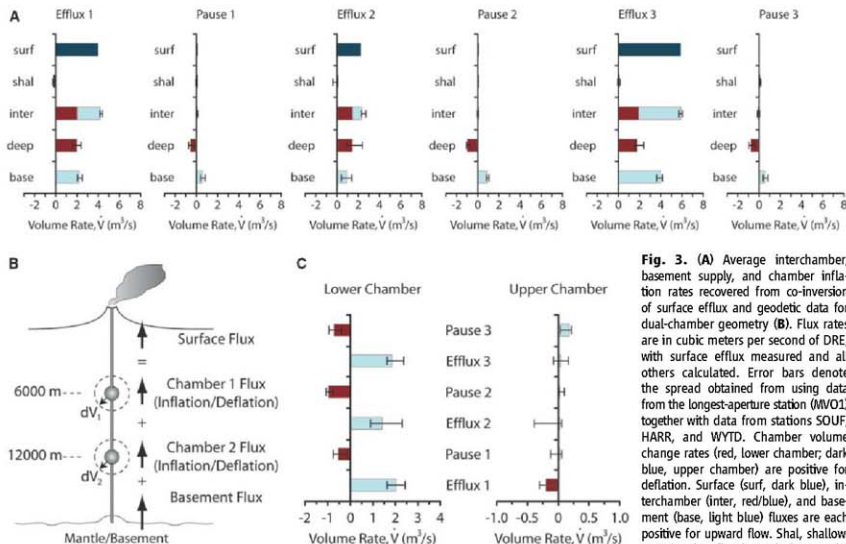
Fig. 3. **(A)** Average interchamber, basement supply, and chamber inflation rates recovered from co-inversion of surface efflux and geodetic data for dual-chamber geometry (**B**). Flux rates are in cubic meters per second of DRE, with surface efflux measured and all others calculated. Error bars denote the spread obtained from using data from the longest-aperture station (MVO1) together with data from stations SOUF, HARR, and WYTD. Chamber volume change rates (red, lower chamber; dark blue, upper chamber) are positive for deflation. Surface (surf, dark blue), interchamber (inter, red/blue), and basement (base, light blue) fluxes are each positive for upward flow. Shal, shallow. Interchamber flux is equivalent to the sum of lower chamber deflation and basement supply (which passes through the lower chamber). Surface efflux is the sum of upper chamber deflation and interchamber transfer (pass-through). **(C)** Inflation (negative, red) and deflation (positive, blue) rates for each of the upper and lower chambers throughout the three sequences of eruption, with a pause after each one.

Fig. 4. Cumulative volume changes for the lower reservoir evaluated from the combined geodetic and efflux histories. E, eruption; P, pause. Modeled depletion is for a spherical chamber within an infinite elastic medium discharging through a vertical conduit. Parameters are as defined in table S2 (chamber radius, 1 km; depth, 12 km; conduit diameter, 30 m; shear modulus of rock, 3 GPa; bulk modulus of magma, 1.1 GPa; magma viscosity 55×10^6 Pa·s). $(B t / \bar{C}) e^{-1}$ represents the characteristic time for chamber depletion controlled by magma and chamber compressibility ($\bar{C}$) and efflux resistance (B).
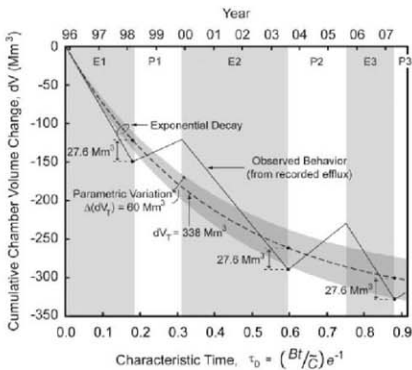
were stable at pressures of ~130 MPa. The presence of small amounts of basalt mixed in the erupted andesite, however, implies that there is a deeper supply of hot mafic magma [10, 11]; some crystal phases also suggest that the upper chamber is connected to a deep reservoir at depths of >10 km [12], and post-1997 geodetic data also support a source possibly as deep as 12 km [3]. The substantial cumulative volume of the eruption (~0.9 km^3) and its decade-long continuity and chemical consistency, coupled with observations of co-eruptive displacements, suggest that the upper magma source is voluminous: ~4 km^3 [13]. Together, these observations constrain a model of two stacked magma reservoirs, at depths of ~6 and ~12 km, connected from the surface to the deep crust and mantle by vertical conduits. Although some evidence suggests that the shallow conduits may be planar [2, 14], the observed surface deformation was radial around the volcano throughout most of the 1997–2007 period (Fig. 1). Correspondingly, the volumetric response of the reservoirs to inflation and deflation appears to dominate the far-field geodetic response, and only these are considered here.

For this model of two stacked magma reservoirs (fig. S1), we co-invert surface efflux and GPS station velocities to recover rates of crustal magma transfer throughout the 12-year duration of the eruption. The three eruptive episodes had sustained surface fluxes of 2 to 8 m^3/s, punctuated by periods of repose with flux <<1 m^3/s (Fig. 2 and Fig. 3A). Over this period, radial

and vertical ground surface velocities are continuously available for at least four GPS stations (Fig. 2). Stations typically show net outward- and upward-deformation during repose and inward- and downward-deformation during renewed surface efflux. Displacement histories are converted to mean surface velocities within each effusive episode or pause (table S1). Mean surface velocities for each of two independent stations (table S1) are then co-inverted with surface efflux [supporting online material (SOM)] to calculate magma migration rates through the intermediate and deep crust (SOM). The mean velocity of the most distal station (MVO1 at ~11 km) (Fig. 1) is combined sequentially with each of the three proximal stations to recover three independent measurements of migration rates and fluxes, shown in Fig. 3 for each of three cycles of eruption followed by a pause (SOM). These results for an incompressible magma (Fig. 3) are representative of calculations for compressible magma for the chamber depths examined here (figs. S2 and S3).

For each of the three eruptive episodes, the co-inversion of the geodetic and efflux data shows that (Fig. 3) the surface efflux responds to volume and pressure changes at a deep level, rather than being a result simply of deflation of a shallow reservoir, as usually presumed. This is apparent in Fig. 3A as an increased magma supply from the basement into the lower chamber, coupled with an active deflation of the lower chamber. For our two-chamber model, the additive flux from these two deeper sources issues into, and causes outflow from, the shallower chamber and upper magmatic system with little volume loss, comprising almost the entire surface efflux in all three active episodes. The only apparent volumetric loss between the deep magmatic system and the surface is manifested as calculated minor inflations (episode 1) or deflations (episodes 2 and 3) of the upper reservoir. Although the upper reservoir is not actively involved as a dynamic storage element, the deeper-sourced material is unlikely to directly transit the upper chamber during a single eruptive episode. The volume change of a spherical reservoir (dV) is proportional to both the pressure change in the magma (dp) and chamber volume (V) and inversely proportional to rock shear modulus (G_R) as $dV \sim (V/G_R)dp$ (SOM eq. 3). For a shear modulus of 1 GPa, co-eruptive inflationary and deflationary volumes of the upper reservoir on the order of 10×10^6 m^3 imply pressure changes of <3 MPa for chamber volumes >4 km^3. For similar moduli and chamber volumes in the lower reservoir, co-eruptive volume changes are an order of magnitude larger (~150×10^6 m^3) and imply co-eruptive pressure drops that are an order of magnitude greater. Correspondingly, the small deformational signal from the upper reservoir implies that either pressure changes are small and the upper system is largely open, or the upper reservoir is smaller and more geometrically rigid in

comparison to a more voluminous lower chamber. Independent geodetic and petrologic evidence identifying the substantial volume of the upper chamber (>4 km^3) favor the presence of an open system.

For the most vigorous of the active phases (episodes 1 and 3, with fluxes for some periods exceeding 7 m^3/s; Fig. 3), the basement flux was larger than the contribution supplied by deflation of the lower reservoir. If the antecedent eruption was particularly vigorous, and therefore substantially depleted the lower reservoir, then this trend was reversed, and the relatively weaker episode 2 was primarily sustained by draining the lower chamber to force out shallower magma. Indeed, for the weak episode 2 (with surface flux steady at 2.2 m^3/s) (Fig. 3), the basal supply was indistinguishable from either the previous or subsequent periods of repose, both of which had fluxes of roughly 1 m^3/s.

During the eruptive pauses, the co-inverted data imply that supply of magma into the basal crust continued at ~1 m^3/s, and the lower reservoir reinflated. Although the lower system appears to have been recharging itself for the next eruptive episode, the upper and lower magmatic systems remained connected, and surface efflux continued at about the same rate of combined supply between the lower and upper reservoirs: Inter-reservoir transfer approximately mirrored the surface efflux (~0.1 to 1.0 m^3/s). Again, the absence of a significant inflationary or deflationary signal in a voluminous upper reservoir as the lower chamber either fills (inflates) or discharges (deflates) suggests that the upper system is open and cannot sustain substantial overpressures transmitted from below. Valving of the flow system between the upper and lower reservoirs is consistent with this observation that the lower reservoir can refill while the upper system remains open. This valving must prevent substantial influx of magma from the lower to the upper chamber during periods of pause, and its charge of either heat or gas, in driving the invigorated system.

With the change from pauses to eruptive phases, magma supply to the deep reservoir continued at a minimum rate of ~1 m^3/s, augmented to rates of ~5 m^3/s. The co-inverted data indicate that the eruptive episodes deplete the lower reservoir only, and not the upper reservoir, which may even inflate slightly as inflow slightly outpaces outflow. During subsequent periods of pause, the deep reservoir reinflates, but typically at half the rate of its previous depletion. Because periods of repose were typically shorter than the periods of active depletion (eruption), the deep reservoir was being depleted (deflated) throughout this decade-long episode. The cumulative volume change for the deep reservoir is illustrated in Fig. 4, indicating that over 12 years, the lower chamber has deflated stepwise by ~320×10^6 m^3, while the upper reservoir inflated by 8×10^6 m^3 (constituting an inflation of 14×10^6 m^3 up to 1998, followed by de-

flation by 6×10^6 m^3 over the remaining decade). This net deflation of the system of ~320×10^6 m^3 is about one-third of the total effusion of ~0.9 km^3 recorded to date, requiring that the remainder of the magma (~570×10^6 m^3) has been sourced from below the lower reservoir. These observations may be compared with models that represent the efflux history from a deflating spherical chamber in an elastic medium (SOM), as illustrated in Fig. 4. This matches the average deflationary history, as shown, and yields a predicted ultimate eruptive volume of 338×10^6 m^3 from the lower chamber with ~320×10^6 m^3 (~95%) transferred to March 2007 (Fig. 4). Although the upper reservoir has been interpreted to be voluminous, on the order of a few cubic kilometers, it is apparent that the major changes in magma storage that have supplied the eruption are from depth (>12 km), with the lower reservoir contributing only a third of the erupted volume.

References and Notes

1. T. H. Druitt, P. Kokelaar, Eds., *The Eruption of Soufrière Hills Volcano, from 1995 to 1999* (Memoir 21, Geological Society of London, London, 2002).
2. G. S. Mattioli *et al.*, *Geophys. Res. Lett.* **25**, 3417 (1998).
3. G. S. Mattioli, R. Herd, *Seismol. Res. Lett.* **74**, 230 (2003).
4. S. R. Young *et al.*, *Geophys. Res. Lett.* **25**, 3389 (1998).
5. G. Norton *et al.*, in (1), pp. 467–482.
6. Montserrat Volcano Observatory, reports from 1995 to 2008, available at www.mvo.ms.
7. W. P. Aspinall *et al.*, *Geophys. Res. Lett.* **25**, 3397 (1998).
8. J. Barclay *et al.*, *Geophys. Res. Lett.* **25**, 3437 (1998).
9. B. Voight *et al.*, *Science* **283**, 1138 (1999).
10. C. Annen, J. D. Blundy, R. S. J. Sparks, *J. Petrol.* **47**, 505 (2006).
11. M. D. Murphy, R. S. J. Sparks, J. Barclay, M. R. Carroll, T. S. Brewer, *J. Petrol.* **41**, 21 (2000).
12. J. D. Devine, M. J. Rutherford, G. E. Norton, S. R. Young, *J. Petrol.* **44**, 1375 (2003).
13. B. Voight *et al.*, *Geophys. Res. Lett.* **33**, L03312 (2006).
14. A. Costa, O. Melnik, R. S. J. Sparks, B. Voight, *Geophys. Res. Lett.* **34**, L02313 (2007).
15. R. S. J. Sparks *et al.*, *Geophys. Res. Lett.* **25**, 3421 (1998).
16. We thank M. Strutt, G. Ryan, and P. Williams of the Montserrat Volcano Observatory and W. Johnston for maintenance of the cGPS network on the SHV. This work is a result of partial support from NSF under awards NSF-CMS-9908590, NSF-CD-0607691, NSF-CD-0507334, and NSF-CD-0607782. The Caribbean Andesitic Lava Island Precision Seismo-geodetic Observatory (CALIPSO) Facility was supported by grants NSF-IF-0523097 and NSF-IF-0732728, and this support is gratefully acknowledged. The conclusions reported here are those of the authors.

Laboratory Simulation of Volcano Seismicity

Philip M Benson,[1,2]* Sergio Vinciguerra,[3] Philip G Meredith,[1] R Paul Young[2]

The physical processes generating seismicity within volcanic edifices are highly complex and not fully understood. We report results from a laboratory experiment in which basalt from Mount Etna volcano (Italy) was deformed and fractured. The experiment was monitored with an array of transducers around the sample to permit full-waveform capture, location, and analysis of microseismic events. Rapid post-failure decompression of the water-filled pore volume and damage zone triggered many low-frequency events, analogous to volcanic long-period seismicity. The low frequencies were associated with pore fluid decompression and were located in the damage zone in the fractured sample; these events exhibited a weak component of shear (double-couple) slip, consistent with fluid-driven events occurring beneath active volcanoes.
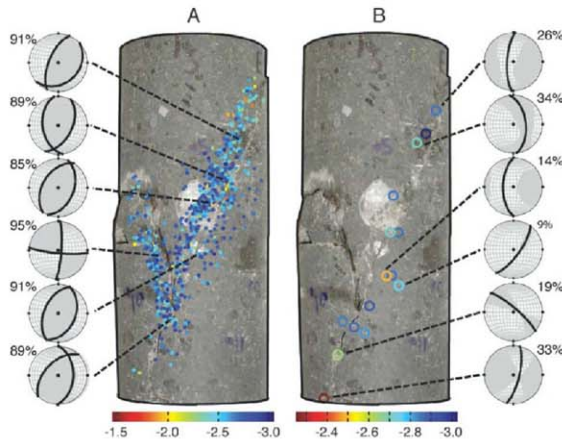
Seismicity in volcanoes is intimately related to ground deformation, faulting, and the movement of volcanic fluids within the edifice. The discrimination and interpretation of different types of seismic signals are central goals for hazard evaluation and mitigation (1–3). Volcanotectonic (VT) seismicity occurs as Earth's crust is deformed and fractured by magma movement, resulting in shallow-focus earthquakes beneath the volcanic edifice, which frequently provide the first sign of growing unrest (4–6). Faulting provides conduits between the deep reservoir and the surface (7), and the final approach to eruption is commonly preceded by accelerating rates in the occurrence of low-magnitude VT earthquakes (5). A second type of seismic activity is related to fluid movement, although the details remain uncertain. Low-frequency (LF) events, also known as long-period and very long-period events, are observed on all types of active volcanoes, often in swarms preceding eruption. These LF events differ from VT events in frequency range and harmonic signature (1, 2, 4) and are postulated to be generated from fluid flow and resonance in fractures and conduits within the edifice. When LF signals become continuous or quasi-continuous, volcanic tremor is generated. It has been suggested that LF earthquakes could provide the basis for a method of eruption forecasting (8). However, because other tectonic processes often yield seismic signatures resembling LF tremor [for example, episodic tremor and slip in subducting regions (9, 10)], further work is urgently needed on investigating the physical processes responsible for LF event generation in volcanic settings.

Here we report results from one of a series of laboratory deformation experiments on Etna basalt samples (11), in which we have generated microseismic signals, better known as acoustic emissions (AEs) (12), that exhibit features nearly identical to those observed during volcanic activity,

when scaled to seismic frequencies. We deformed cylindrical samples of Etna basalt (50 mm in diameter and 125 mm long, with a centrally predrilled 3-mm-diameter conduit) at an effective confining pressure of 60 MPa, representative of pressure conditions approximately 2 km beneath volcanic edifices (13, 14). The initial pore fluid pressure was 20 MPa, giving an effective confining pressure of 40 MPa. Experiments were conducted in two stages. In stage 1, samples were deformed at a constant strain rate of $4 \times 10^{-6} \ s^{-1}$ until brittle failure occurred. This resulted in the creation of a localized shear fault and an associated crack

damage zone. In stage 2, the fluid stored in the sample (including both preexisting microcracks and the new fault damage zone) was rapidly decompressed from the top of the sample (15) (fig. S1), stimulating rapid fluid flow out of the sample via the damage zone and the pre-drilled conduit. The output of AEs was recorded continuously throughout both phases with an array of 16 transducers. In addition to recording waveforms, the array was used to locate individual AE events by using a downhill simplex routine and a triaxial velocity model (11, 15). We estimate that our AE locations are accurate to ±1.5 mm (11). We also computed source characteristics (mechanisms) of the located events via relative amplitude moment tensor analysis (16, 17). Not only does our experiment concur with field data in which fluid movement is frequently a key mechanism behind seismic swarms in general (18) [in addition to being ubiquitous in volcanic areas, where it is often implicated in rapid volume change processes such as fluid motion in conduits associated with magma decompression (8)], but we were also able to resolve moment tensors for independent verification of pore volume changes that are a feature of fluid-driven earthquakes.

We observed a low rate of AE output during initial loading, as preexisting cracks closed. Activity then increased exponentially as differential

[1]Rock and Ice Physics Laboratory, Department of Earth Sciences, University College London, Gower Street, London, WC1E 6BT, UK. [2]Lassonde Institute, University of Toronto, 170 College Street, Toronto, Ontario, M5S 3E3, Canada. [3]Istituto Nazionale di Geofisica e Vulcanologia, Sezione di Roma 1, Via di Vigna Murata 605, 00143, Rome, Italy.

*To whom correspondence should be addressed. E-mail: p.benson@ucl.ac.uk

Fig. 1. Post-test sample of Etna basalt showing a single throughgoing fault and its conjugate. Superimposed are AE locations resulting from the deformation phase of the experiment (A) (dots) and AE locations resulting from the rapid decompression of the pore fluid (B) (open circles). The color bar indicates the dimensionless event pseudo-magnitude, calculated using an average weighted ray path over all receivers (19). Irrespective of the stage of the experiment (deformation or decompression), events are located on or close to the fault. Source characteristics (mechanisms) associated with deformation (A) exhibit high percentage components of DC shear; whereas those relating to pore fluid decompression (B) show much lower DC components.

stress increased and new, dilatant cracks nucleated and propagated (11, 15, 17, 19). Failure occurred by shear faulting at a failure stress of 480 MPa and 2.9% axial strain (15) (fig. S2). We recorded 1518 events during deformation and located 762. Of these, most were located within the damage zone of the shear fault or its conjugate (Fig. 1A). They exhibited dominantly double-couple (DC) source characteristics (Fig. 1A), entirely as expected for the faulting phase of the experiment. We verified that the presence of the central conduit did not substantially affect the mechanics of deformation and failure by conducting identical experiments without a conduit and obtaining essentially identical results.

The rapid decompression of the pressurized pore fluid (over 0.2 s) by means of a valve at the top of the apparatus (15) (fig. S1) was accompanied by a swarm of AE events, again located near the fault damage zone generated during stage 1 (Fig. 1B and fig. S3). In contrast with the deformation phase of the experiment, the source

Fig. 2. Representative waveform types observed during the experiment. (**A**) High-frequency events analogous to VT seismicity are observed during sample deformation and fracture formation. These signals have a maximum power component at 600 kHz (and above), with less power at lower frequencies. (**B**) LF events observed during pore space decompression. These signals have a peak power component at approximately 18 to 20 kHz. Referring to the normalized power plots, for VT seismicity (A) there is a rapid onset of high-frequency power with a peak at ~0.14 ms and little power lower frequencies; the high frequencies die out rapidly. However, for LF frequencies, there is no high-frequency component to speak of, and the LF component dies very slowly, with a maximum power occurring at ~0.6 ms.

Fig. 3. FESEM micrograph taken along the center axis of the sample. (**Left**) AE locations from the fracture zone decompression stage (stage 2) of the experiment. Locations are sparse because of the low-amplitude signals, but enough data are located to show the trend of locations within the fracture damage zone. (**Top right**) Some additional example waveforms, with their corresponding locations. (**Bottom right**) Details of the local microstructure of two areas (A and B) where clusters of LF events occurred, showing voids, undulating cracks walls, and pinch points in the damage zone.
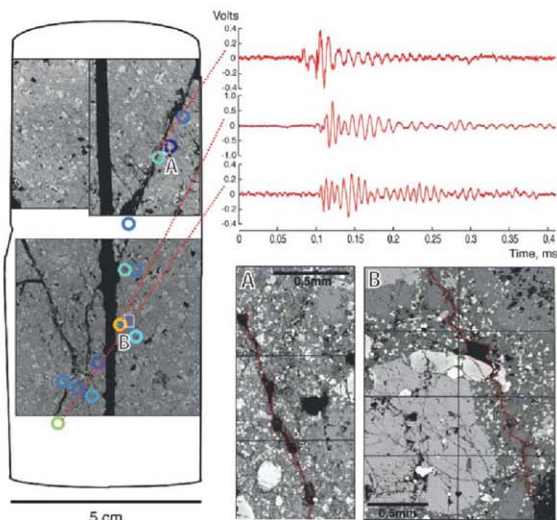
characteristic of the decompression-related events exhibited low components of shear but a high volumetric component (Fig. 1B). We therefore postulate that the rapid flow of the pore fluid through a tortuous fracture damage zone, imposed by the rapid decompression, created conditions conducive for the formation of a swarm of AE events analogous to LF events recorded at active volcanoes. Source mechanisms involving high levels of volumetric change have been widely reported in volcanic areas (13, 20) and in areas of tectonic subduction and fault overpressure (18, 20), all of which have been linked to fluid movement.

Further similarities between field data and our laboratory data can be seen in the frequency domain (Fig. 2). During the rock deformation stage, high-frequency AE signals dominated as the shear fault nucleated and propagated across the sample (Fig. 1A). Brittle processes, from millimeter-scale laboratory microcracking to kilometer-scale earthquake rupture, generate high-frequency movement and acceleration. In our experiment, the peak in power occurred at a frequency of approximately 600 kHz (Fig. 2A) during sample deformation and fracture. A characteristic feature of volcano seismic events due to fluid movement is their LF signature. This type of LF event has principal power components of 1 to 2 Hz (4, 12), and most authors postulate that resonance within cracks or volcanic conduit flow is responsible for such seismicity, which often shows non-DC source characteristics (10, 20, 21). In our experiments,

the decompression of the fluid-filled pore space resulted in swarms of AE activity with a power frequency of approximately 18 to 50 kHz (Fig. 2B). The events had noticeably longer, ringing waveforms than those from deformation events, but with noticeable harmonic overtones—features commonly seen in field seismic data (1, 4). Experimental studies using open-ended tubes of both constant and changing section have revealed similar frequency characteristics (22, 23). In addition to these two key end-member waveform types, our experiments yielded a high-frequency initial component as well as LF characteristic that had features of both VT and LF event types. We postulate that these events occurred when the fluid-filled rock faulted and fluids moved into the fracture. Such waveforms are commonly known as hybrid events in volcano seismology (4).

To investigate the microstructural origin of the located AE events, we analyzed field emission scanning electron microscopy (FESEM) images of the deformed and decompressed sample. As shown in Fig. 3, a complex damage zone formed in the lower half of the sample, dominated by two major conjugate faults. During decompression, events were located within this damage zone. The detailed locations of two particular AE clusters imply that the fluids producing these events were following highly tortuous pathways, with many pinch-outs and undulating features. Such geometries have long been postulated to be responsible for tremor-type

events (24). FESEM observations showed that many of the cracks were filled with broken and comminuted rock (Fig. 3, inset B), as also reported from field observations of fractured magma and obsidian (25). Taken together, these observations suggest that LF events in volcanic areas are generated when hydrothermal fluids (water, steam, dusty gases, and/or magma itself) move through preexisting crack networks (4, 13) comprising both larger faults and their associated fracture damage zones (Figs. 3 and 4).

Using a simple size-frequency scaling relationship (13, 26), we show that LF events with laboratory length (d_L) and frequency (f_L) scales (50 mm and 18 to 50 kHz, respectively) can be scaled appropriately to data from source dimensions (d_V) and seismic frequencies (f_V) typical of natural volcanic events (200 m to 1 km and 1 to 2 Hz, respectively) (6). AE obeys power-law relationships, just as field scale seismicity does, permitting similar statistics to be used (27, 28). After the treatment of (13) and (26), one may write $d_L \times f_L = d_V \times f_V$, if the two processes show the same scale characteristics. With the measurements quoted above, this yields $d_V/d_L = 4 \times 10^3$ to 20×10^3 and $f_L/f_V = 9 \times 10^3$ to 50×10^3, which is in excellent agreement. Although somewhat simplistic, this first-order treatment confirms that our laboratory data scale to natural volcanic data. Likewise, other parameters relevant to volcano physics, such as viscosity (V), can also be scaled in this simple way. Using $V_L = 10^{-3}$ Pa·s for laboratory pore water and $V_V = 10^2$ Pa·s for basaltic lava, respectively, and length scales of $d_L = 50$ mm and $d_V = 10$ km (common for effusive eruptions), we find that $d_V/d_L = 2 \times 10^5$ and $V_V/V_L = 1 \times 10^5$, again providing excellent agreement.

We further propose that the recorded LF AE events acted as individual trigger sites in our sample for the longer LF tail. This hypothesis is supported by the locations of individual LF events, which are not confined to a particular site within the sample damage zone, indicating different sources within the sample. The source of the LF tremor induced by the trigger swarm is therefore likely to be related to the conduit itself (8). This can be verified using the dimensions of the laboratory setup and the simple relationship velocity = $f \lambda$. For the known conduit length (λ) of 125 mm and a compressional wave velocity ranging from 1500 m/s (water) to 6000 m/s (unfractured basalt), frequencies (f) are between 12 and 48 kHz. This is in good agreement with the recorded resonances of approximately 20 to 50 kHz. Our laboratory data therefore support field evidence (1, 18, 25, 29), high-temperature laboratory evidence from partial melts (26), and model evidence (4, 8, 13) that LF events are generated as a combined result of fluid interactions with both the damage zone and the conduit.
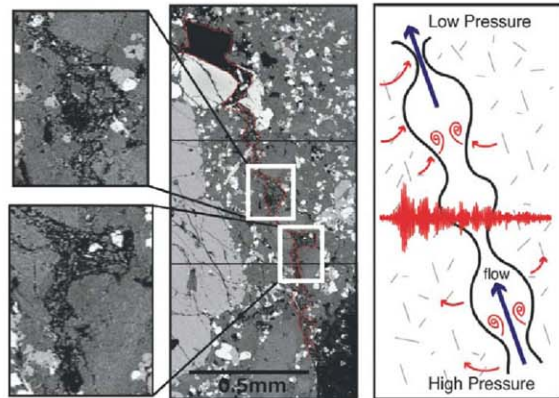
Fig. 4. (Left and center) Post-test laboratory FESEM micrographs showing the key features of a typical undulating crack structure; kinks, constrictions, and comminuted fill material. **(Right)** Conceptual model of damage zone geometry based on laboratory evidence and the tremor model of Julian (24), with additional features taken from the aseismic plug flow model of Neuberg (8). We also observe comminuted rock fragments within the damage zone, analogous to ash in fractures as observed in the field by Tuffen and Dingwell (25) and hypothesized by Chouet (1). It is likely that any nonlaminar flow produced by the undulations and kinks in the damage zone is recorded as LF seismicity due to conduit/crack resonance, as these features are commonly seen at the location of the LF AE events in our experiment.

References and Notes
1. B. A. Chouet, *Nature* **380**, 309 (1996).
2. H. Kumagai et al., *Science* **293**, 687 (2001).
3. B. Voight et al., *Science* **283**, 1138 (1999).
4. B. Chouet, *Pure Appl. Geophys.* **160**, 739 (2003).
5. C. R. J. Kilburn, *J. Volcanol. Geotherm. Res.* **125**, 271 (2003).

6. S. R. McNutt, in *Monitoring and Mitigation of Volcanic Hazards*, R. Scarpa, R. Tilling, Eds. (Springer Verlag, Berlin, 1996), pp. 99–146.
7. R. Scandone, K. V. Cashman, S. D. Malone, *Earth Planet. Sci. Lett.* **253**, 513 (2007).
8. J. W. Neuberg *et al.*, *Geotherm. Res.* **153**, 37 (2006).
9. G. Rogers, H. Dragert, *Science* **300**, 1942 (2003).
10. Y. Ito *et al.*, *Science* **315**, 503 (2007).
11. P. M. Benson *et al.*, *Geophys. Res. Lett.* **34**, 10.1029/2006GL028721 (2007).
12. AE has its origin in mechanical engineering and high-pressure hydraulics, where it is frequently used to detect leaks from gas tanks (such as liquified petroleum gas flashing to gas). The turbulent flow generates AE.
13. K. Aki, P. G. Richards, *Quantitative Seismology* (University Science Books, Sausalito, CA, ed. 2, 2002).
14. D. Patane *et al.*, *Science* **299**, 2061 (2003).
15. P. M. Benson, S. Vinciguerra, P. G. Meredith, R. P. Young, paper presented at the 29th Course of the International School of Geophysics Euro-Conference of Rock Physics and Geomechanics, Erice, Sicily, 28 September, 2007.
16. D. C. Collins, W. S. Pettitt, R. P. Young, *Pure Appl. Geophys.* **159**, 197 (2002).
17. S. Stanchits *et al.*, *Pure Appl. Geophys.* **163**, 975 (2006).
18. S. Miller *et al.*, *Nature* **427**, 10.1038/nature02251 (2004).
19. D. A. Lockner *et al.*, *Nature* **350**, 39 (1991).
20. G. R. Foulger *et al.*, *Geotherm. Res.* **132**, 45 (2004).
21. H. Kumagai, B. A. Chouet, M. Nakano, *J. Geophys. Res.* **107**, 10.1029/2001JB001704 (2002).
22. M. R. James *et al.*, *Geotherm. Res.* **129**, 61 (2004).
23. M. R. James *et al.*, *J. Geophys. Res.* **111**, 10.1029/2005JB003718 (2006).
24. B. R. Julian, *J. Geophys. Res.* **99**, 10.1029/93JB03129 (1994).
25. H. Tuffen, D. Dingwell, *Bull. Volcanol.* **67**, 370 (2005).
26. L. Burtini *et al.*, *Geology* **35**, 183 (2007).
27. I. G. Main, *Nature* **357**, 27 (1992).
28. C. G. Hatton, I. G. Main, P. G. Meredith, *Nature* **367**, 160 (1994).
29. C. A. Rowe *et al.*, *Geophys. Res. Lett.* **25**, 2297 (1998).
30. This work was supported by a Marie-Curie International Fellowship within the 6th European Community Framework program (contract MOIF-CT-2005-020167 to P.M.B.), project FIRB-MIUR (Sviluppo Nuove Tecnologie per la Protezione e Difesa del Territorio dai Rischi Naturali) to S.V.; and a Canadian Foundation for Innovation award to R.P.Y. The authors thank L. Malagnini and J. Neuberg for fruitful discussions, A. Cavallo for FESEM technical support at the HT-HP laboratory (INGV Rome), and two anonymous reviewers for comments that greatly helped to improve this manuscript.

Northern Hemisphere Controls on Tropical Southeast African Climate During the Past 60,000 Years

Jessica E. Tierney,[1]* James M. Russell,[1] Yongsong Huang,[1] Jaap S. Sinninghe Damsté,[2] Ellen C. Hopmans,[2] Andrew S. Cohen[3]

The processes that control climate in the tropics are poorly understood. We applied compound-specific hydrogen isotopes (δD) and the TEX$_{86}$ (tetraether index of 86 carbon atoms) temperature proxy to sediment cores from Lake Tanganyika to independently reconstruct precipitation and temperature variations during the past 60,000 years. Tanganyika temperatures follow Northern Hemisphere insolation and indicate that warming in tropical southeast Africa during the last glacial termination began to increase ~3000 years before atmospheric carbon dioxide concentrations. δD data show that this region experienced abrupt changes in hydrology coeval with orbital and millennial-scale events recorded in Northern Hemisphere monsoonal climate records. This implies that precipitation in tropical southeast Africa is more strongly controlled by changes in Indian Ocean sea surface temperatures and the winter Indian monsoon than by migration of the Intertropical Convergence Zone.

The mechanisms that cause fluctuations of rainfall and temperature in the tropics—home to a large portion of the world's population and a region of central importance to the global hydrologic cycle—are poorly understood. One process often invoked to explain past changes in tropical precipitation and temperature is a shift in the mean annual position of the Intertropical Convergence Zone (ITCZ), which migrates meridionally in response to seasonally and orbitally driven changes in interhemispheric heat distribution (1, 2). However, loci of tropical precipitation and convergence within the ITCZ itself respond to changes in sea surface temperatures (SSTs) and coupled ocean-atmosphere zonal modes, such as El Niño–Southern Oscillation and the Indian Ocean Dipole (IOD) (3, 4), as well as the strength of the monsoons that bring moisture into the continents (5). The importance of these zonal forces acting on tropical rainfall, vis-à-vis changes in ITCZ position, is incompletely understood, in part because long, high-resolution paleoclimatic reconstructions that constrain tropical rainfall and temperature are rare.

Fig. 1. Map of East Africa and Lake Tanganyika. Two Kullenberg piston cores collected from the Kalya Horst (S 6°42′, E 29°50′) in 2004 (cores NP04-KH04-3A-1K and NP04-KH04-4A-1K) were selected to compile a continuous sedimentary record for the past 60,000 years. The Tanganyika basin and watershed span 2° to 10°S, covering much of the southeast African tropics. See SOM text 1 for regional climatology.

Tropical Africa is a site of deep tropical convection and contains a population that is economically vulnerable to climatic change. The African Great Lakes, situated in the East African Rift Valley, are excellent archives of climate change in tropical East Africa. Reconstructions of lake level fluctuations establish that East Africa was generally dry during the Last Glacial Maximum (LGM), humid during the early Holocene, and dry during the late Holocene (6). Furthermore, high-resolution records from these lakes based on indicators such as fossil pollen, geochemistry, and lithology show that East African environmental changes are linked to global climate processes on millennial time scales (7–11). Yet available records from East Africa rarely extend beyond the LGM and do not independently constrain rainfall and temperature. This has confounded the interpretation of humid and arid intervals in this region; the Younger Dryas, for example, is variously implicated to be wet and dry in tropical southeast Africa (8, 11).

To better understand the controls on East African temperature and hydrology, we applied the TEX$_{86}$ (tetraether index of 86 carbon atoms) paleotemperature proxy together with hydrogen isotopic analyses of terrestrial leaf waxes to sediment cores from Lake Tanganyika, at an average time step of 300 years (Fig. 1). TEX$_{86}$ is based on

[1]Department of Geological Sciences, Brown University, Providence, RI 02912, USA. [2]Department of Marine Organic Biogeochemistry, Royal Netherlands Institute for Sea Research (NIOZ), 1790 AB Den Burg, Netherlands. [3]Department of Geosciences, University of Arizona, Tucson, AZ 85721, USA.

*To whom correspondence should be addressed. E-mail: Jessica_Tierney@brown.edu

the relative degree of cyclization of aquatic archaeal glycerol dialkyl glycerol tetraether (GDGT) isoprenoidal lipids; is linearly correlated with temperature (12); and can be applied to reconstruct the temperature of some large lakes, including Lake Tanganyika (13, 14) [supporting online material (SOM) text 2].

Terrestrial plants are the dominant source of long-chain (C_{26} to C_{30}) monocarboxylic fatty acids in lacustrine sediments, and their deuterium/hydrogen (D/H) ratio is an excellent indicator of terrestrial hydrologic conditions (15, 16). We analyzed the δD of the C_{28} n-acid [hereafter $\delta D_{leaf\ wax}$, expressed in per mil versus Vienna standard mean ocean water (SMOW) values], the most abundant fatty acid in our cores. Because the isotopic fractionation that occurs during leaf wax synthesis appears consistent across different plant types (due to the interactive effects of vegetation type and relative humidity, $\delta D_{leaf\ wax}$ reflects changes in the δD of precipitation (15). The primary control on δD of tropical East African precipitation is the "amount effect" (17), although moisture source and history may play a secondary role. Thus, higher δD values indicate reduced precipitation, and lower values represent wet periods.

Our TEX_{86} and $\delta D_{leaf\ wax}$ reconstructions show that temperature and hydrology in the Tanganyika basin were extremely variable throughout the past 60,000 years (Fig. 2). Holocene lake surface temperature (LST) fluctuated between ~27° and 29°C, whereas temperatures during the LGM were ~4°C cooler. The magnitude and timing of this temperature shift are similar to those of nearby Lake Malawi (14), indicating that our TEX_{86}

record captures regional temperature change in tropical southeast Africa during deglaciation.

$\delta D_{leaf\ wax}$ spans a range of ~50 per mil (‰,) and records numerous abrupt shifts between arid and humid conditions throughout the past 60,000 years. During marine isotope stage (MIS) 3, rapid excursions toward enriched $\delta D_{leaf\ wax}$ values indicate millennial-scale pulses of aridity, the most pronounced of which is centered at 37,255 years before the present (yr B.P.) ± 917 yr B.P. This event is coincident with Heinrich event 4 (H4), which occurred at 38,000 yr B.P. and is the largest of the six Heinrich events during MIS 3 in the North Atlantic (18). It is possible that arid events at 47,500 and 57,000 yr B.P. in our record are associated with H5 and H6 (Fig. 3), although our age control is not sufficient to constrain their timing. A positive 15‰ shift in $\delta D_{leaf\ wax}$ at ~16,700 yr B.P. indicates that the most arid conditions of the past 60,000 years occurred during H1, yet we do not observe changes in $\delta D_{leaf\ wax}$ during H2 and H3. The dramatic expression of H4 and H1 in our record and the lack of H2 and H3 suggest a variable sensitivity of climate in the Tanganyika basin to North Atlantic climate processes, possibly due to the different magnitudes of the Heinrich events themselves and their ability to propagate into the southern tropics.

Climate in the Tanganyika basin switched from arid conditions during the LGM and H1 to humid conditions in the early Holocene in two rapid steps. A 260-year-long 17‰ decrease in $\delta D_{leaf\ wax}$ occurs at 15,100 yr B.P., coincident with the onset of the Bølling period (19), followed by a 270-year-long 23‰ decrease at

11,600 yr B.P., the end of the Younger Dryas (19) (Fig. 3). The early Holocene was the wettest period in our record, as indicated by $\delta D_{leaf\ wax}$ values that are depleted by 30‰ relative to late Holocene values. The transition toward a more arid late Holocene at about 4,700 yr B.P. was also abrupt, occurring within 300 years. Such rapid changes from humid to arid conditions are a persistent feature of our $\delta D_{leaf\ wax}$ record, suggesting that rainfall in this region responds in a nonlinear fashion to changing climatic parameters.

Tanganyika LST did not cool during the millennial-scale arid intervals recorded by $\delta D_{leaf\ wax}$ (Fig. 3). However, precipitation and temperature do co-vary on orbital time scales (Fig. 2). Shifts between warm/wet and cool/dry conditions follow Northern Hemisphere summer insolation, as opposed to austral summer or annual insolation at 6°S, indicating that the heat and moisture budget of this part of Africa is dynamically linked to the Northern Hemisphere.

In particular, Tanganyika LST at the end of the LGM follows rising Northern Hemisphere summer insolation, a potential trigger for deglaciation (20). Temperatures rise at 20,000 ± 380 yr B.P., just as they do in a TEX_{86} LST record from Lake Malawi (14) (Fig. 3). This timing is consistent with rising temperatures at ~20,000 yr B.P. in Antarctica, yet leads the deglacial CO_2 rise recorded in Antarctic ice cores (21) by about 3,000 years, a difference that is outside the chronological errors of the ice core and the LST records. Increasing greenhouse gas concentrations are therefore not responsible for the initial transmission of warming from the high latitudes to the southeast African tropics. Yet a rapid, presumably atmospheric, communication mechanism must exist for this region to "feel" Northern Hemisphere insolation. Aside from CO_2, changes in the hydrologic cycle are suspect, but our $\delta D_{leaf\ wax}$ record indicates that precipitation did not increase until 15,100 yr B.P. Thus, the mechanism linking deglacial temperature changes between high and low latitudes remains elusive.

Our $\delta D_{leaf\ wax}$ data are remarkably similar to isotopic records of the Asian monsoon from Hulu and Dongge Caves in China (22, 23), with arid events occurring during the Younger Dryas, H1, and H4 to H6; a wet early Holocene; and a dry late Holocene, although the timing of their respective transitions to a more arid late Holocene is different (Fig. 3). This in-phase behavior is surprising, because Northern and Southern Hemisphere rainfall records should be out of phase if ITCZ position were the dominant control on rainfall variability in the Tanganyika basin. During the early Holocene "African Humid Period," for example, the ITCZ was situated farther north over Africa (24), and the region around Lake Tanganyika should have become more arid as the ITCZ spent less time in the southern tropics. Likewise, we observe that the Younger Dryas was arid in the Tanganyika basin, though a southward shift of the ITCZ would predict humid conditions (2, 9). Proxy evidence from Lake Malawi also indicates

Fig. 2. $\delta D_{leaf\ wax}$ and TEX_{86} LST from Lake Tanganyika. $\delta D_{leaf\ wax}$ and TEX_{86} data from core NP04-KH04-3A-1K are plotted in blue and red, and data from core NP04-KH04-4A-1K are plotted in purple and orange. The TEX_{86} data show a 0.5° offset between the two cores; this is probably due to core NP04-KH04-4A-1K's more proximal location to the eastern shore of the lake, where seasonal coastal upwelling occurs today (SOM text 3). $\delta D_{leaf\ wax}$ data line up well between cores, reflecting the regional character of the proxy. Chronology is constrained by 33 ^{14}C AMS dates, which are plotted for NP04-KH04-3A-1K and

NP04-KH04-4A-1K as black and orange triangles, respectively (SOM materials and methods). June-July-August (JJA) insolation for 30°N (solid line) emphasizes the Northern Hemisphere influence evident in the temperature and precipitation data. December-January-February (DJF) insolation for the local area (6°S) (dotted line) is shown for contrast.
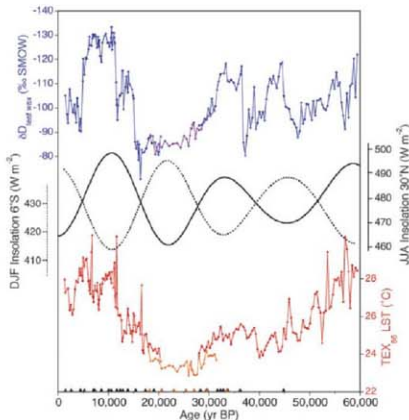
that the Younger Dryas and H1 to H5 were dry [10, 11], implying that precipitation patterns over much of tropical southeast Africa did not respond predictably to ITCZ migration. This behavior is distinctly different from the rainfall history of southern tropical South America, which is clearly anti-phased with the Northern Hemisphere [25].

If ITCZ position alone cannot explain rainfall variability in tropical southeast Africa, factors controlling the advection of moisture to the ITCZ, as well as the strength of convergence itself, must be implicated. Indeed, SSTs, which control latent heat flux into the atmosphere; ocean-basin SST gradients, which influence large-scale atmospheric circulation patterns; and the strength of the winds that advect humidity into the continents are all important to tropical African precipitation [26, 27]. These factors could effect changes in precipitation in the Lake Tanganyika basin via several different mechanisms. Wet conditions could arise from an increase in Atlantic Ocean moisture flux via the Congo basin, as proposed by Schefuß et al. [27] to explain humid conditions during the early Holocene in central Africa. However, moisture flux into the Congo basin is low at ~11,000 yr B.P. and does not reach a maximum until ~8,000 yr B.P. [27], which is substantially different from the timing of changes in $\delta D_{leaf\,wax}$ that we observe in Lake Tanganyika.

Alternatively, we argue that changes in moisture flux into Africa from the Indian Ocean basin control East African precipitation variability. Modern precipitation in southeast Africa is highly seasonal, occurring from October to April, and most of this moisture is derived from the Indian Ocean [28]. Thus, we expect precipitation to vary in concert with Indian Ocean SSTs (which govern the generation of moisture), as well as the strength of the winter Indian monsoon circulation (which affects the transport of moisture). Although the winter and summer monsoons are commonly thought to work in opposition to one another [29], global climate modeling experiments show that on orbital time scales, increases in Northern Hemisphere precession result in amplified seasonality that should intensify both the summer and winter monsoons [30, 31]. A more vigorous winter Indian monsoon, then, could explain the humid conditions that prevailed in the Tanganyika basin during the Early Holocene as well as during precession maxima in MIS 3.

Winter monsoon strength does not, however, satisfactorily explain the arid episodes we observe during the Younger Dryas, H1, and H4.

Zr/Ti data from Lake Malawi indicate strong northerly winds in southeast Africa during these events [10], implying vigorous winter monsoon circulation and a southward migration of the ITCZ. Yet these events are distinctly arid in both the Tanganyika and Malawi basins. We therefore surmise that cold SSTs during millennial-scale stadials reduced latent heat flux from the western Indian Ocean, minimizing the amount of moisture transported into East Africa by the winter monsoon. Data from the Pakistan margin that indicate cooler SSTs during millennial-scale stadials in MIS 3 support this hypothesis [32].

Thus the mechanisms controlling precipitation in tropical southeast Africa evidently vary over different time scales, though they are not necessarily mutually exclusive; Indian Ocean SSTs may also play a role in influencing orbital-scale changes in East African rainfall. During the early Holocene, enhanced cooling in the eastern Indian Ocean is thought to have modified Walker circulation over the basin, producing longer and more frequent positive IOD events [30]. The positive IOD phase is associated with warmer SSTs in the western Indian Ocean, an intensification of convergence, and an increase in East African rainfall [4]. This mechanism, along with an increase in the Indian seasonal monsoons, could account for the humid conditions that characterize tropical East Africa during this time.

The fact that both temperature and precipitation in the Tanganyika basin show many characteristics of Northern Hemisphere climatic variability demonstrates that the Northern Hemisphere has a substantial influence on climate in tropical southeast Africa. An early warming during the last deglaciation highlights the importance of atmospheric teleconnections linking the high and low latitudes, although the mechanisms behind this relationship remain unknown. Although ITCZ migration strongly influences East African climate by modifying seasonal wind strength [9, 10], it is apparently not the dominant control on precipitation variability in tropical southeast Africa; rather, oceanic and atmospheric factors controlling moisture generation and advection heavily influence rainfall variability in this region. Particularly, the highly nonlinear character of our $\delta D_{leaf\,wax}$ record implies that precipitation regimes in East Africa are liable to change abruptly in response to climatic forcing. This behavior has implications for modeling the response of East African hydrology to anthropogenic climate change.

Fig. 3. Comparison of Lake Tanganyika TEX$_{86}$ and $\delta D_{leaf\,wax}$ data with other records of paleoclimate. Records are plotted as follows: (**A**) $\delta^{18}O$ data from Hulu and Dongge caves [22, 23]. (**B**) Lake Tanganyika $\delta D_{leaf\,wax}$. (**C**) Deglacial CO_2 record from the Epica Dome C ice core [21]. (**D**) Lake Malawi TEX$_{86}$ LST [14]. (**E**) Lake Tanganyika TEX$_{86}$ LST, with the NP04-KH04-4A-1K values corrected for the 0.5°C offset described in Fig. 2. Gray bars indicate the Younger Dryas (YD) and H1 to H6 as recorded in Hulu Cave and Lake Tanganyika. The onset of warming in East Africa is marked with a dotted vertical line.
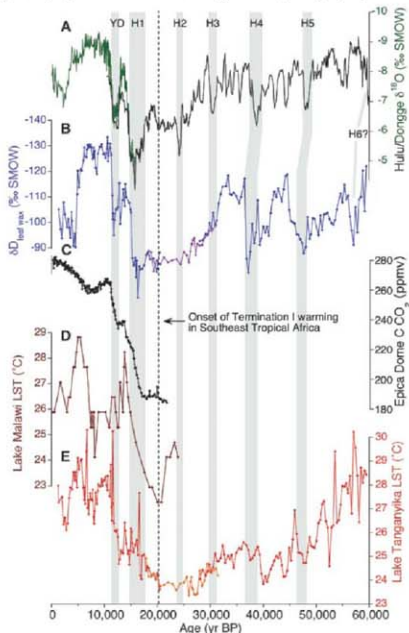
References and Notes

1. G. H. Haug, K. A. Hughen, D. M. Sigman, L. C. Peterson, U. Röhl, *Science* **293**, 1304 (2001).
2. D. W. Lea, D. K. Pak, L. C. Peterson, K. A. Hughen, *Science* **301**, 1361 (2003).
3. C. M. Moy, G. O. Seltzer, D. T. Rodbell, D. M. Anderson, *Nature* **420**, 162 (2002).
4. N. H. Saji, B. N. Goswami, P. N. Vinayachandran, T. Yamagata, *Nature* **401**, 360 (1999).
5. P. J. Webster *et al.*, *J. Geophys. Res.* **103**, 14451 (1998).
6. F. Gasse, *Quat. Sci. Rev.* **19**, 189 (2000).
7. R. Bonnefille, F. Chalié, *Global Planet. Change* **26**, 25 (2000).

8. Y. Garcin, A. Vincens, D. Williamson, J. Guiot, G. Buchet, *Geophys. Res. Lett.* **33**, 10.1029/2005GL025531 (2006).
9. J. E. Tierney, J. M. Russell, *Geophys. Res. Lett.* **34**, 10.1029/2007GL029508 (2007).
10. E. T. Brown, T. C. Johnson, C. A. Scholz, A. S. Cohen, J. W. King, *Geophys. Res. Lett.* **34**, 10.1029/2007GL031240 (2007).
11. I. S. Castañeda, J. P. Werne, T. C. Johnson, *Geology* **35**, 823 (2007).
12. J. Kim, S. Schouten, E. C. Hopmans, B. Donner, J. S. Sinninghe Damsté, *Geochim. Cosmochim. Acta* **72**, 1154 (2008).
13. L. A. Powers *et al.*, *Geology* **32**, 613 (2004).
14. L. A. Powers *et al.*, *Geophys. Res. Lett.* **32**, 10.1029/2004GL022014 (2005).
15. J. Hou, W. D'Andrea, Y. Huang, *Geochim. Cosmochim. Acta* **72**, 3503 (2008).
16. B. Shuman, Y. Huang, P. Newby, Y. Wang, *Quat. Sci. Rev.* **25**, 2992 (2006).
17. M. Vuille, M. Werner, R. S. Bradley, F. Keimig, *J. Geophys. Res.* **110**, 10.1029/2005JD006022 (2005).
18. S. R. Hemming, *Rev. Geophys.* **42**, 1 (2004).
19. R. B. Alley, P. U. Clark, *Annu. Rev. Earth Planet. Sci.* **27**, 149 (1999).
20. K. Kawamura *et al.*, *Nature* **448**, 912 (2007).
21. E. Monnin *et al.*, *Science* **291**, 112 (2001).
22. Y. J. Wang *et al.*, *Science* **294**, 2345 (2001).
23. D. Yuan *et al.*, *Science* **304**, 575 (2004).
24. P. deMenocal *et al.*, *Quat. Sci. Rev.* **19**, 347 (2000).
25. X. Wang *et al.*, *Geophys. Res. Lett.* **34**, 10.1029/2007GL031149 (2007).
26. P. Barker, F. Gasse, *Quat. Sci. Rev.* **22**, 823 (2003).
27. E. Schefuß, S. Schouten, R. R. Schneider, *Nature* **437**, 1003 (2005).
28. Y. Zhu, R. E. Newell, *Mon. Weather Rev.* **126**, 725 (1998).
29. G. Yancheva *et al.*, *Nature* **445**, 74 (2007).
30. N. J. Abram *et al.*, *Nature* **445**, 299 (2007).
31. A. B. G. Bush, *Global Planet. Change* **32**, 331 (2002).
32. M. J. Higginson, M. A. Altabet, L. Wincze, T. D. Herbert, D. W. Murray, *Paleoceanography* **19**, 10.1029/2004PA001031 (2004).
33. We thank two anonymous reviewers for their insightful comments; S. Schouten, S. Clemens, and T. Herbert for their suggestions on earlier drafts of the manuscript; and M. Alexandre, N. Meyer, J. Ossebaar, and I. Castañeda for analytical assistance. This research was supported by NSF-EAR 0639474 to J.R., the Nyanza Project (grants NSF-ATM 0223920 and BIO 0383765, to A.C.), and the National Defense Science and Engineering Graduate Fellowship to J.T. The authors declare that they have no competing financial interests.

Natural Selection on a Major Armor Gene in Threespine Stickleback

Rowan D. H. Barrett,* Sean M. Rogers, Dolph Schluter

Experimental estimates of the effects of selection on genes determining adaptive traits add to our understanding of the mechanisms of evolution. We measured selection on genotypes of the *Ectodysplasin* locus, which underlie differences in lateral plates in threespine stickleback fish. A derived allele (low) causing reduced plate number has been fixed repeatedly after marine stickleback colonized freshwater from the sea, where the ancestral allele (complete) predominates. We transplanted marine stickleback carrying both alleles to freshwater ponds and tracked genotype frequencies over a generation. The low allele increased in frequency once lateral plates developed, most likely via a growth advantage. Opposing selection at the larval stage and changing dominance for fitness throughout life suggest either that the gene affects additional traits undergoing selection or that linked loci also are affecting fitness.

Adaptive evolution occurs when genetic variation affects phenotypes under selection. This process has been detected by the discovery of candidate genes underlying phenotypic traits whose adaptive significance is known or suspected (*1–7*) and by identifying statistical signatures of selection on genomic regions affecting phenotypic traits (*8–12*). However, field experiments evaluating the fitness consequences of allelic substitutions at candidate loci should provide estimates of the timing and strength of selection, enhance understanding of the genetics of adaptation, and yield insights into the mechanisms driving changes in gene frequency.

Freshwater threespine sticklebacks (*Gasterosteus aculeatus*) originated from marine populations that invaded newly created coastal lakes and streams throughout the Northern Hemisphere following the last ice age. Within the past 20,000 years or less, freshwater populations repeatedly underwent a loss in bony armor plating (*13*). Marine sticklebacks are typically armored with a continuous row of 30 to 36 bony lateral plates on each side (complete morph), whereas freshwater sticklebacks typically have 0 to 9 plates (low morph) or, less often, an intermediate number of plates (partial morph) (*13–15*) (Fig. 1). Armor reduction following colonization of freshwater evolved rapidly (*16–19*) from the fixation of a clade of low alleles of the *Ectodysplasin* gene (hereafter, the *Eda* low allele). This allele evolved ~2 million years ago and is rare (~1%) in the ocean (*1*). The repeated fixation of this allele implies that it undergoes positive selection in freshwater, because genetic drift alone is unlikely to produce a strong correlation between phenotype and environment (*20–22*).

Fish with reduced armor have a juvenile growth advantage (*23*), which may result from the higher cost of mineralizing bone in freshwater (*24, 25*), which has low ion concentrations relative to marine environments. This increased growth rate should, in turn, reduce predation by insects (*26*), as well as increase lipid stores, which results in higher over-winter survival (*27*). Larger fish also may breed earlier (*28*) and have access to better terri-

Fig. 1. Lateral plate morphs in marine stickleback. Complete morph (top), partial morph (middle), and low morph (bottom). Fish were stained with Alizarin red to highlight bone.

10 mm

Zoology Department and Biodiversity Research Centre, University of British Columbia, 6270 University Boulevard, Vancouver, BC V6T 1Z4, Canada.

*To whom correspondence should be addressed. E-mail: rbarrett@zoology.ubc.ca

tories, an increase in mating success, and a higher reproductive output (28–36). To test this hypothesis, we tracked adaptive evolution at the *Eda* locus in replicated transplants of marine stickleback to freshwater environments. We predicted that we would observe positive selection on the low allele via advantages in growth, survival, and reproduction. We also looked for deviations from this expectation, which might suggest that *Eda* or linked genes have unexpected fitness effects.

We experimentally introduced adult wild marine fish heterozygous at the *Eda* locus to four freshwater ponds (37). The fish were trapped from a marine stickleback population in southwestern British Columbia. We introduced approximately equal numbers of these fish ($n = 45$ to 46) to each pond in the spring of 2006, initiating replicate freshwater invasions. Within 60 days, we observed larval fish in each colonized pond, indicating that the marine colonizers were breeding. Genotyping of four microsatellite markers, which were all in linkage equilibrium with *Eda*, confirmed that nearly all alleles present in the parents were at similar frequencies in the progeny (fig. S1), which suggested that founding events did not saturate our sampling artifacts. Genotype frequencies at *Eda* in the F_1 generation were not significantly different from the predicted 1:2:1 ratio (Fig. 2A) [pond 1: $\chi^2(2) = 0.06$, $P = 0.97$; pond 2: $\chi^2(2) = 1.09$, $P = 0.58$; pond 3: $\chi^2(2) = 1.09$, $P = 0.58$; and pond 4: $\chi^2(2) = 1.20$, $P = 0.55$]. Subsequently, we sampled 50 fish from each pond 10 times over 1 year to monitor changes in offspring allele frequencies.

We observed strong fluctuations in *Eda* allele and genotype frequencies, with replicate ponds showing nearly parallel oscillations (Fig. 2A). We did not observe strong changes in allele frequency in the unlinked microsatellite markers, which suggested that these results are not due to demographic effects (fig. S1). Fish achieved their adult number of lateral plates after reaching a standard length of ~30 mm (25, 38, 39). Most experimental fish passed this threshold between October and November 2006 [average length in October was 27.32 mm (± 5.99 SD); average length in November was 33.14 mm (± 4.70 SD)]. In agreement with our predictions for growth, by October, juvenile fish carrying the low allele were larger than juvenile fish homozygous for the complete allele. Mean body length was positively associated with the number of low alleles per genotype in all ponds [one-tailed t test of four slopes, $t(3) = 2.53$, $P = 0.043$]. We also noted higher overwintering survival rates in fish with the low allele. From October 2006 to May 2007, the frequency of the complete allele dropped from 67 to 49%, which reflected the comparatively poor survival of individuals homozygous for the complete allele. We calculated that the selection coefficient (S) against the complete allele between these dates was 0.52 (± 0.10 SEM) (Fig. 2) (37).

At the start of the breeding season in May 2007, the number of low alleles carried by an individual was again positively associated with body length in all ponds [one-tailed t test of four slopes, $t(3) = 2.35$, $P = 0.050$], and sexually mature individuals were significantly larger than nonbreeding individuals (Fig. 3) [Welch two-tailed t tests, pond 1: $t(6) = 2.47$, $P = 0.049$; pond 2: $t(2) = 9.40$, $P = 0.006$; pond 3: $t(9) = 2.61$, $P = 0.027$; and pond 4: $t(13) = 4.23$, $P < 0.001$]. The genotypes of the earliest reproductive individuals were biased toward carrying the low allele compared with nonreproductive individuals, with 95% being heterozygote or homozygous low (Fig. 3) [tested by the interaction between breeding status and genotype in a log-linear model, $\chi^2(2) = 7.30$, $P = 0.026$; no effects of pond were detected, $\chi^2(6) = 2.88$, $P = 0.82$]. By July 2007, most individuals had reached sexual maturity, and we observed little difference in genotype frequencies between sexually mature individuals and the overall population (Fig. 3) [$\chi^2(2) = 2.56$, $P = 0.28$]. By this time, we also could not detect a correlation between size and *Eda*

genotype [$t(3) = -0.30$, $P = 0.607$]. In all four ponds, the frequency of the low allele was greater in the first sample of F_2 offspring in June 2007 than in all F_1 adults sampled in May 2007 [June F_2: 57.0% (± 4.1% SEM), May F_1: 51.6% (± 1.4% SEM)] (Fig. 2A) [one-tailed t test, $t(3) = 2.14$, $P = 0.061$]. By July, the frequency of the low allele in F_2 juveniles had decreased to 52.2% (± 3.7% SEM), which reflected the similar genotypic ratios of breeding and nonbreeding adults later in the breeding season.

These patterns linking the low *Eda* allele with higher growth, improved survival, and earlier breeding are consistent with the hypothesis that positive selection stemmed from a reduced burden of producing armor plates in freshwater. This effect, combined with the possibility of reduced vertebrate predation pressure in freshwater compared with the sea (25, 40), may account for the evolution of low genotype populations with reduced plates in freshwater. At the same time, selection against plate production does not fully

Fig. 2. (A) Frequency of the low allele in four replicate ponds (different colored lines). All samples are from the first (F_1) cohort of offspring, except the June and July 2007 samples, which are from the second (F_2) pond generation. **(B)** Approximate life history stages through the course of the experiment. Fish stained as in Fig. 1. **(C)** Genotype frequencies averaged across all four ponds. All samples are as in (A). Purple, homozygous complete genotype (CC); orange, heterozygote genotype (CL); green, homozygous low genotype (LL). Vertical bars show standard errors on the basis of $n = 4$ ponds.
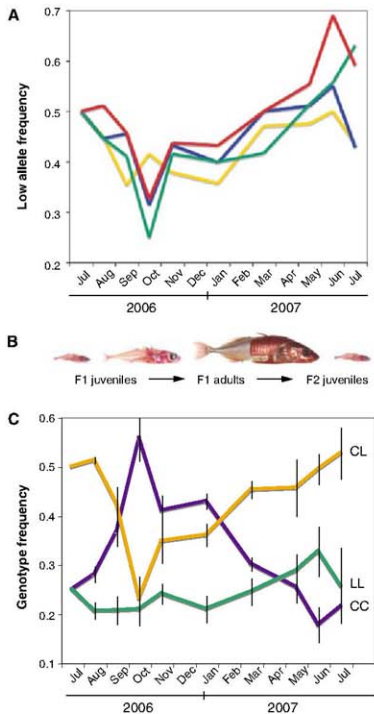
explain the observed changes in *Eda* allele frequencies. We noted selection favoring the complete allele in all four ponds (Fig. 2A) very early in life, before the fish attain the size at which number of lateral plates is finalized (about 30 mm). The calculated selection coefficient (S) against the low allele between July and October 2006 was 0.50 (± 0.16 SEM) (Fig. 2C), which offset the gains occurring later in life. We also observed oscillations in the relative fitness of heterozygotes at *Eda*, which are difficult to explain solely in terms of the burden of lateral plates, because the size and number of plates in heterozygotes are intermediate between low and complete homozygotes (22). The decline in low *Eda* allele frequencies early in life was associated with a drop in the frequency of heterozygous fish and a rise in the frequency of the homozygous complete genotype, which suggested that there is heterozygote underdominance for fitness at this stage [$h = -1.38$ (± 0.23 SEM)]. Underdominance was especially apparent by October 2006, when heterozygous fish made up less than 25% of the total in our samples, instead of the 50% observed at the start of the F_1 cohort. This episode was followed by a period between November 2006 and May 2007 during which the heterozygotes at *Eda* had the highest fitness of all three genotypes [$h = 2.57$ (± 0.98 SEM)]. Although positive selection favored the low allele during this period, heterozygotes increased in frequency much faster than the homozygous low genotype (Fig. 2C). These findings suggest

that either variation at the *Eda* gene has direct or epistatic effects on other phenotypic traits contributing to fitness, or it is linked to another, unidentified locus affecting fitness.

Our results highlight the utility of direct measurements of natural selection on genes for understanding the genetic basis of adaptation by enabling us to test a mechanism favoring reduction of lateral plates in freshwater environments. Many of our results are consistent with selection against high plate number, although they do not rule out the possibility that selection is also occurring on genes tightly linked to *Eda* (1). Our results also expose opposing selection on *Eda* early in life similar in magnitude to the measured advantage of the low allele later in life. This demonstrates not only that countervailing selection pressures diminish the advantage of the low allele over the whole life span but also that the overall fitness effects of *Eda* do not seem to be determined solely by differences in lateral plate number. Along with the fluctuating dominance in fitness at the *Eda* locus, these results indicate that there may be additional pleiotropic effects of this gene. This work underscores the need for a synthesis of population biology and genomics, to determine the genetic basis of fitness differences in natural populations (41).

References and Notes

1. P. F. Colosimo *et al.*, *Science* 307, 1928 (2005).
2. A. Abzhanov, M. Protas, B. R. Grant, P. R. Grant, C. J. Tabin, *Science* 305, 1462 (2004).
3. R. C. Albertson, J. T. Streelman, T. D. Kocher, P. C. Yelick, *Proc. Natl. Acad. Sci. U.S.A.* 102, 16287 (2005).
4. H. D. Bradshaw, K. G. Otto, B. E. Frewen, J. K. McKay, D. W. Schemske, *Genetics* 149, 367 (1998).
5. H. E. Hoekstra, R. J. Hirschmann, R. A. Bundey, P. A. Insel, J. P. Crossland, *Science* 313, 101 (2006).
6. M. D. Shapiro *et al.*, *Nature* 428, 717 (2004).
7. S. M. Rogers, I. Bernatchez, *Mol. Biol. Evol.* 24, 1423 (2007).
8. J. M. Akey *et al.*, *PLoS Biol.* 2, e286 (2004).
9. R. Nielsen *et al.*, *PLoS Biol.* 3, e170 (2005).
10. J. A. Shapiro *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 104, 2271 (2007).
11. J. C. Wootton *et al.*, *Nature* 418, 320 (2002).
12. S. I. Wright *et al.*, *Science* 308, 1310 (2005).
13. M. A. Bell, S. A. Foster, *The Evolutionary Biology of the Threespine Stickleback* (Oxford Univ. Press, Oxford, 1994).
14. M. A. Bell, *Copeia* 1977, 277 (1977).
15. D. W. Hagen, L. G. Gilbertson, *Heredity* 30, 273 (1973).
16. T. Klepaker, *Can. J. Zool.* 71, 1251 (1993).
17. B. K. Kristjansson, S. Skulason, D. L. G. Noakes, *Evol. Ecol. Res.* 4, 659 (2002).
18. B. K. Kristjansson, *Environ. Biol. Fishes* 74, 357 (2005).
19. M. A. Bell, W. E. Aguirre, N. J. Buck, *Evolution Int. J. Org. Evolution* 58, 814 (2004).
20. G. G. Simpson, *The Major Features of Evolution* (Columbia Univ. Press, New York, 1953).
21. H. D. Rundle, L. Nagel, J. W. Boughman, D. Schluter, *Science* 287, 306 (2000).
22. D. Schluter, E. A. Clifford, M. Nemethy, J. S. McKinnon, *Am. Nat.* 163, 809 (2004).
23. K. B. Marchinko, D. Schluter, *Evolution Int. J. Org. Evolution* 61, 1084 (2007).
24. N. Giles, *J. Zool.* 199, 535 (1983).
25. M. A. Bell, G. Orti, J. A. Walker, J. P. Koenings, *Evolution Int. J. Org. Evolution* 47, 906 (1993).
26. S. A. Foster, V. B. Garcia, M. Y. Town, *Oecologia* 74, 577 (1988).
27. R. A. Curry, S. L. Currie, S. K. Arndt, A. T. Bielak, *Environ. Biol. Fishes* 72, 111 (2005).
28. E. T. Schultz, L. M. Clifton, R. R. Warner, *Am. Nat.* 138, 1408 (1991).
29. U. Candolin, H. R. Voigt, *Evolution Int. J. Org. Evolution* 57, 862 (2003).
30. S. Einum, I. A. Fleming, *Evolution Int. J. Org. Evolution* 54, 628 (2000).
31. D. Hasselquist, *Ecology* 79, 2376 (1998).
32. A. Aebischer, N. Perrin, M. Krieg, J. Studer, D. R. Meyer, *J. Avian Biol.* 27, 143 (1996).
33. A. P. Møller, *Behav. Ecol. Sociobiol.* 35, 115 (1994).
34. L. Rowe, D. Ludwig, D. Schluter, *Am. Nat.* 143, 698 (1994).
35. K. Landa, *Evolution Int. J. Org. Evolution* 46, 121 (1992).
36. S. Verhulst, J. M. Tinbergen, *J. Anim. Ecol.* 60, 269 (1991).
37. Materials and methods are available as supporting material on *Science* Online.
38. M. A. Bell, *Genetica* 112–113, 445 (2001).
39. M. A. Bell, *Evolution Int. J. Org. Evolution* 38, 665 (1981).
40. T. E. Reimchen, *Behaviour* 137, 1081 (2000).
41. H. Ellegren, B. C. Sheldon, *Nature* 452, 169 (2008).
42. We thank K. Marchinko, A. Paccard, M. Dosani, K. Dhahan, J.-B. Oboni, E. Sciampe, C. Verchere, J. Courchesne, E. J. Davis, F. Guillaume, J. Hill, C. Jordan, C. Spencer, M. Barrueto, M. Dionne, and S. Rogers for field and genotyping assistance and S. Otto, R. Grant, S. Barrett, and the Schluter lab group for comments on the manuscript. Supported by Natural Sciences and Engineering Research Council of Canada (NSERC Discovery and Special Research Opportunity grants (D.S.), grants from the Canada Foundation for Innovation (D.S.), and NSERC postdoctoral (S.M.R.) and graduate scholarships (R.D.H.B.). Microsatellite sequences were deposited in GenBank by the Stanford Genome Research Center with the accessions BV678144, BV678119, BV678140, and BV67814.
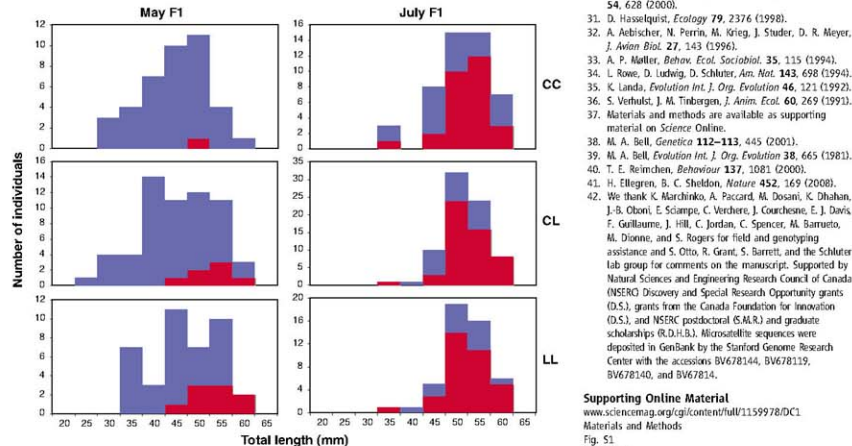
Fig. 3. Body length of individuals in the first (F_1) pond cohort during the breeding season, in May and July 2007, summed across all ponds. Red, individuals in reproductive condition; blue, individuals not in reproductive condition. *Eda* genotypes are labeled on the right axis: homozygous complete (CC), heterozygous (CL), and homozygous low (LL).

257

Global Warming, Elevational Range Shifts, and Lowland Biotic Attrition in the Wet Tropics

Robert K. Colwell,[1]* Gunnar Brehm,[2]† Catherine L. Cardelús,[3]†
Alex C. Gilman,[4]† John T. Longino[5]†

Many studies suggest that global warming is driving species ranges poleward and toward higher elevations at temperate latitudes, but evidence for range shifts is scarce for the tropics, where the shallow latitudinal temperature gradient makes upslope shifts more likely than poleward shifts. Based on new data for plants and insects on an elevational transect in Costa Rica, we assess the potential for lowland biotic attrition, range-shift gaps, and mountaintop extinctions under projected warming. We conclude that tropical lowland biotas may face a level of net lowland biotic attrition without parallel at higher latitudes (where range shifts may be compensated for by species from lower latitudes) and that a high proportion of tropical species soon faces gaps between current and projected elevational ranges.

R ecent global climate change has already begun to affect species' geographic ranges. Poleward shifts in range limits correlated with warming climate and changes in precipitation have been documented for a wide spectrum of temperate and subtropical species, and phenological changes portend poleward shifts in even greater numbers (1). Corresponding upslope shifts in range boundaries along temperate elevational gradients have also been detected for both plants and animals (1–3).

In contrast, no study has documented climate-driven, contemporary latitudinal range shifts in the tropics. Biogeographical baseline data and resurveys at the high resolution required to document incipient range shifts are still scarce for tropical regions. More fundamentally, however, the steady latitudinal temperature gradients that c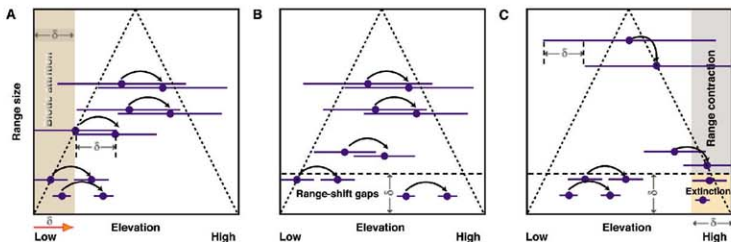haracterize temperate latitudes level off to a broad plateau within the tropics (fig. S1). For midtropical species of restricted geographical range and limited capability for rapid range shift, the shallow latitudinal temperature gradient within the tropics presents a daunting obstacle to poleward range shifts (4).

In the tropics as elsewhere, however, elevational temperature gradients are steep—from 5.2°C to 6.5°C decrease per 1000 m elevation (5)—nearly 1000 times as much as the latitudinal rate of decrease in temperature, per km, in the temperate zone (6.9° per 1000 km at 45° N or S) (fig. S1) and vastly greater than the modest latitudinal temperature gradient within the tropics. Consequently, for tropical species affected by warming climates, upslope range shifts appear far more likely than latitudinal shifts (4, 6). In fact, elevational shifts provide the only published evidence, to date, for contemporary range shifts within the tropics—a handful of examples, all for vertebrates, from Monteverde, Costa Rica (7).

To illustrate the potential for elevational range shifts in the tropics, we analyze elevational range data for four large survey data sets of plants and insects [epiphytes (8), understory Rubiaceae (5), geometrid moths (9), and ants (5)]. The data for all 1902 species were collected by the authors since 2001 from the Barva Transect, a continuously forested corridor ascending 2900 m up an elevational gradient from La Selva Biological Station, near sea level, to the top of Volcán Barva, in Costa Rica (8) (fig. S2). Before turning to an analysis of these data, however, we first examine the challenges that tropical species may face on such a gradient.

At temperate latitudes, resources released by poleward or upslope range shifts may be appropriated by species from lower latitudes or lower elevations—species already adapted to warmer temperatures—if suitable habitat corridors and dispersal mechanisms permit (1). Likewise, on tropical mountainsides, upslope range shifts may be compensated for by the influx of species currently found at lower elevations or by expansion from small nuclei left over from previous warming episodes (4, 6). In the tropical lowlands, however, the parallels end. No community of species, now living in even hotter places, is available to replace tropical lowland species

[1]Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT 06269, USA. [2]Institut für Spezielle Zoologie und Evolutionsbiologie mit Phyletischem Museum, Friedrich-Schiller-Universität Jena, Erbertstraße 1, 07743 Jena, Germany. [3]Department of Biology, Colgate University, Hamilton, NY 13346, USA. [4]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90950, USA. [5]Evergreen State College, Olympia, WA 98505, USA.

*To whom correspondence should be addressed. E-mail: colwell@uconn.edu
†These authors are listed alphabetically.

Fig. 1. A simple graphical model of the potential effects of climate warming on the distribution of species ranges on a bounded elevational gradient. The elevational range size for each species (vertical axis) is plotted as a function of its elevational midpoint (horizontal axis), with corresponding range limits indicated by the solid horizontal lines. Because ranges cannot extend beyond the limits of the gradient, all range-size/range-midpoint coordinate pairs lie within the geometric constraint triangle (dotted lines). (The constraint triangle is not to be confused with a mountain.) An upslope shift in isotherms with warming climate is measured (in meters elevation) by δ, the sole parameter of the model. **(A)** Lowland biotic attrition. If lowland ranges shift δ m upslope, their new lower range limits must lie at or above δ m elevation, predicting a decrease in lowland richness. **(B)** Range-shift gaps. If elevational ranges follow isotherms upslope, no projected range smaller than δ m (plotted below the dashed horizontal line) will overlap its prewarming elevational range, posing challenges for dispersal and establishment. **(C)** Range contraction and mountaintop extinction. All ranges with upper limits less than δ m from the upper domain limit are predicted to contract, and the smallest of these (those less than δ m in extent) face local extinction.

that shift upslope with climate change or those that become extinct if no suitable habitat corridor to cooler climates is accessible.

How likely is "lowland biotic attrition," the net loss of species richness in the tropical lowlands from upslope range shifts and lowland extinctions driven by climate warming? Surprisingly, this question has scarcely been considered in a broad biogeographical framework. To explore it, we must ask whether tropical lowland species are already living near the thermal optimum of their climatic fundamental niche, above which fitness would decline in the absence of acclimation or adaptation. For plants, especially, because temperature and precipitation interact strongly through transpirational water loss, the answer depends on future changes in precipitation as well as in temperature. Wallace wrote that "[i]n the equable equatorial zone there is no ... struggle against climate. Every form of vegetation has become alike adapted to its genial heat and ample moisture, which has probably changed little even throughout geological periods" (10). We now know that lowland tropical climates have changed substantially and relentlessly ever since species-rich forests resembling modern ones first

occupied the lowland wet tropics in the mid-Tertiary (6). Although the notion of long-term constancy of tropical climates is now universally dismissed, Wallace's view of tropical climates as benign lingers on (11), underlying the apparently widespread conviction that "[m]any tropical species may well be able to withstand higher temperature[s] than those in which they currently exist" (12).

Global climate has been cooling since the Middle Miocene [14.5 million years before the present (yr B.P.)], a trend punctuated, but not reversed, by the repeated, dramatic temperature fluctuations of the Quaternary glacial cycles (1.8 million yr B.P. to present). Pollen core data and plant microfossils from tropical sites, worldwide, show plant species moving downslope during the Last Glacial Maximum (20,000 yr B.P.), and back up again with the Holocene warming (10,000 yr B.P.) (13, 14). For the Andes, similar records span multiple Pleistocene glaciations (4).

Although the climatic changes driving them were complex (4, 15), these range shifts attest to the sensitivity of montane tropical species to temperature regimes (indeed, it is routine to estimate paleotemperatures from paleovegetation). Are tropical lowland species somehow exceptional,

having retained the tolerance for high-temperature environments of their Tertiary ancestors? If so, that tolerance has survived the adaptive demands of multimillennial glacial episodes when the lowland tropics averaged some 5°C cooler than now (a mean temperature currently characteristic of 30° N or S latitude), with much colder extreme events in some regions (6, 13, 16). Although habitat heterogeneity certainly mitigated extinction (4), lowland species or genotypes especially well adapted to warm climates, which "had nowhere to go" (6) during glacial episodes, would seem the least likely to have survived these cold periods unchanged, because of adaptive tradeoffs (17).

The warm interglacial periods of the Pleistocene, instead of Tertiary climates, might provide more appropriate adaptive benchmarks for warm climate tolerance. Current evidence, however, suggests that contemporary warming, 0.25°C decade^{-1} since 1975 in the tropical lowlands (15), has already driven global mean temperature to within ~1°C of Earth's maximum temperature in the past million years, exceeding the Holocene maximum (9000 to 5000 yr B.P.) of the current interglacial (18). In short, tropical climates are now warmer than at any time during at least the past two million years (4).

These considerations suggest, on evolutionary grounds, that many lowland tropical species may be in for trouble if they do not shift to higher elevations or to cooler, wetter microhabitats in coming decades—and trouble may already be at hand for some. Experimental studies indicate that tropical ectotherms are living at temperatures already near their thermal optimum, with early fitness declines expected under current projections for climate warming (19). The growth rate of individual lowland tropical trees is slowing, correlated on a year-to-year basis with increasing mean annual temperature in Costa Rica (20) (at La Selva) (fig. S2), and plot-level measurements in Panamanian and Malaysian forests show similar trends (21). In the Amazon, elevated atmospheric CO_2 itself may be interacting with increasing temperatures to drive changes in tropical forest composition (22), and distribution models for individual species project widespread range contractions and declines in tree viability and species richness (23), based on current bioclimatic envelopes. Some global climate models suggest that Amazonian forest may be approaching a "critical resiliency threshold," beyond which the system may lose its capacity to counter the effects of warming on tropical plants (24)

Species will respond individualistically to tropical climate change, just as they have in the past (4, 14), and vegetation types may expand or contract (6, 15). Nevertheless, the tropical lowlands are likely to experience decreased species richness, with novel plant communities (25) composed of heat-tolerant species [including fire-adapted, drought-tolerant species in areas subject to decreased precipitation (14, 15, 26)]. Early successional species, adapted to germination at higher soil temperatures, may thrive at the
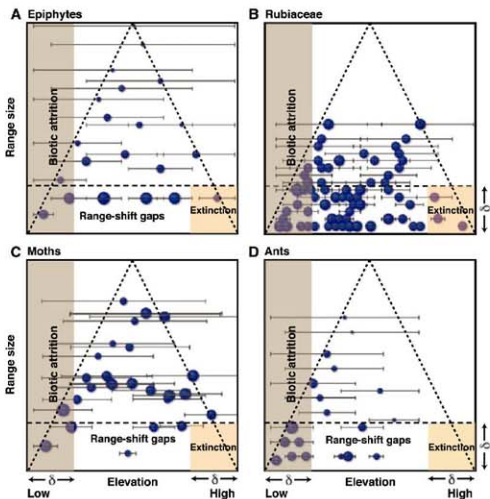
Fig. 2. The model in Fig. 1 applied to four groups of species on the Barva Transect (30 to 2900 m elevation) in Costa Rica (fig. S2). An upslope range shift of 600 m elevation is assumed, based on 3.12°C climate warming [the Intergovernmental Panel on Climate Change median regional rate for the next century (32) is 3.2°C] and locally measured lapse rate (5). Each blue disc represents the midpoint-range coordinate pair for n species (symbol size proportional to $\sqrt{n}$) (see Fig. 1 legend for details). (**A**) Data for 555 species of epiphytes (8) sampled at six elevations. (**B**) Data for 82 species of understory Rubiaceae (5) sampled at 28 elevations. (**C**) Data for 739 species of geometrid moths (9) sampled at six elevations. (**D**) Data for 495 species of ants (5) sampled at six elevations.

expense of late-successional species that require cooler microhabitats. Anthropogenic habitat alteration, shifts in land use patterns, and exotics and invasives (27, 28) can be expected to exacerbate many of these effects (4, 26).

Lowland biotic attrition is not the only biogeographic consequence of warming climate that is likely to be a greater problem in the tropics than in the temperate zone. Gaps between current and projected ranges ("range-shift gaps"), a concern at all latitudes, are especially worrisome on tropical elevational gradients because of the great number of tropical species with narrow elevational ranges. If tropical elevational ranges are narrower than their temperate counterparts [as widely believed, albeit on sparse evidence (29)], then tropical species are more likely than temperate species to experience range-shift gaps for a given upslope shift in climatic isotherms. Species already living near the top of elevational gradients face "mountaintop extinction" unless they have disjunct populations elsewhere on higher mountains or at cooler latitudes (3, 25, 27, 30, 31).

To illustrate the potential for lowland biotic attrition, range-shift gaps, and mountaintop extinctions in the wet tropics, we analyzed data for 1902 species of epiphytes, understory rubiaceous plants, geometrid moths, and ants on the Barva Transect, spanning 2900 m elevation (8) (Fig. S2). Species distribution modeling (e.g., 12, 23, 30, 31), commonly used to project range shifts under climate change, requires data for the full geographical ranges of species and corresponding environmental variables. Unfortunately, such data do not exist for the vast majority of tropical species (including those in this study), with the principal exception of the best-known vertebrate groups (e.g., 31). Instead, we designed a simple graphical model (Fig. 1) that relies on temperature to assess potential elevational range shifts for transect data. In Figs. 2 and 3, we apply the model to illustrate the potential consequences of warming-driven, elevational range shifts for the four groups of organisms surveyed on the Barva Transect.

Because the four groups of species differ substantially in statistical distribution of elevational range extents and in patterns of range location along the elevational gradient (Fig. 2), they project different patterns of potential range-shift effects (Fig. 3) (5). Overall, for a 600-m upslope shift in isotherms, driven by a 3.2°C temperature increase over the next century (32), about half (53%) of the 1902 species in this study (those that currently reach the lowest elevations) are candidates for lowland biotic attrition, and about half (51%) may be faced with range-shift gaps. The potential for mountaintop extinctions for these species on the Barva Transect is minimal for a 600-m shift in isotherms (Fig. 3A) but begins to appear at about a 1000-m range shift (Fig. 3B). Although mountaintop extinction has been the focus of attention in most discussions of elevational range shifts (e.g., 2, 27, 31), in the near term a far greater proportion of the tropical species in this study, and elsewhere, are threatened with lowland attrition or are challenged by early spatial discontinuities between current and projected elevational range (range-shift gaps). Many face both challenges.

In many respects, the predictions illustrated in Fig. 3 must be considered worst-case scenarios, even if warming occurs as assumed. Estimating elevational range limits from local inventory data is likely to underestimate regional elevational range, even accounting for local undersampling (5). Our projections share with species distribution models (12, 23, 30, 31) the assumptions that the fundamental climatic niche of each species is fully expressed by current distributions; that the effects of climate outweigh any idiosyncratic effects of species interactions, dispersal limitation, demographic patterns, or historical contingency; that change will be too rapid for adaptation to warmer temper-
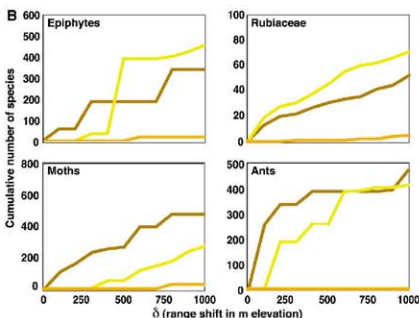
atures at lower range limits (1, 4, 30); and that habitats at the landscape scale are homogeneous with regard to microclimate. In fact, species that currently occupy warmer microhabitats at their lower range limit, including lowland species, may shift to currently cooler (and wetter) refuges at the same elevation, in response to warming (1, 4, 14, 16).

On the other hand, our simple projections based on temperature fail to take account of factors that may exacerbate the challenges facing tropical species. In the lowlands, decreased precipitation and increased fire frequency may amplify the direct effects of increased temperature (15, 26). Anthropogenic habitat fragmentation and widespread interruption of elevational forest corridors, which may increase as cloud levels rise with climate warming (4, 7, 26), cannot fail to impede successful range shifts, particularly for forest-dependent species with limited dispersal potential (4, 14, 26, 28). As current local assemblages are shuffled by individualistic range shifts, key species interactions may be disrupted (1, 4, 7, 28). Narrow-ranged species faced with range-shift gaps will have to compete successfully with wide-ranged species that continue to occupy upslope portions of their current ranges (14). Range-shift projections focus on immediate consequences, ignoring long-term effects of decreased suitable habitat and smaller populations, where land area declines with increasing elevation (27) as it does on the Barva Transect (33). Finally, the current spectrum of climate types may contract or novel types appear, as thermal bands move upslope (25).

Wallace's impression of tropical heat as "genial" may have survived the dismissal of his view of tropical climates as unchanging, but we suggest that it is time to question the conventional wisdom that tropical climates are biologically benign by taking a closer look at the challenges that climate change poses for tropical species. In the tropics, successful latitudinal range shifts ap-
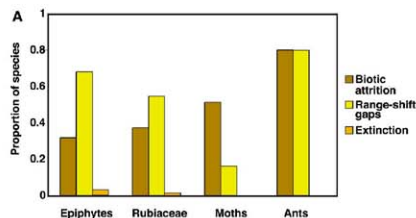
Fig. 3. (A) Proportion of species in each of the four groups of Fig. 2 subject to decline or disappearance in the lowlands (biotic attrition), faced by gaps between current and projected elevational range (range-shift gaps), and exposed to mountaintop extinction, given a 600-m upslope shift in all ranges. Proportions sum to greater than 1 because some species belong in two categories. **(B)** Cumulative number of species facing each of these three challenges as a function of warming-driven range shifts. The x axis represents model parameter δ, measured in meters of elevation range shift, on a continuous scale of warming-driven isotherm shifts of up to 5°C (nearly 1000 m), the upper range of projections for Central America for this century (32). The stairstep patterns are a consequence of sampling at discrete sites on the gradient (5).

pear unlikely, putting the focus on elevational gradients, where range-shift gaps will develop early for the great numbers of narrow-ranged species. The lowland tropics lack a source pool of species adapted to higher temperatures to replace those driven upslope by warming, raising the possibility of substantial attrition in species richness in the tropical lowlands.

References and Notes

1. C. Parmesan, *Annu. Rev. Ecol. Evol. Syst.* **37**, 637 (2006).
2. R. Wilson, D. Gutierrez, J. Gutierrez, V. Monserrat, *Glob. Change Biol.* **13**, 1873 (2007).
3. J. Lenoir, J. Gegout, P. Marquet, P. de Ruffray, H. Brisse, *Science* **320**, 1768 (2008).
4. M. B. Bush, *Glob. Ecol. Biogeogr.* **11**, 463 (2002).
5. Materials and methods are available as supporting material on *Science* Online.
6. M. B. Bush, H. Hooghiemstra, in *Climate Change and Biodiversity*, T. E. Lovejoy, L. Hannah, Eds. (Yale Univ. Press, New Haven, CT, 2005), pp. 125–137.
7. J. A. Pounds, M. P. L. Fogden, K. L. Masters, in *Climate Change and Biodiversity*, T. E. Lovejoy, L. Hanna, Eds. (Yale Univ. Press, New Haven, CT, 2005), pp. 70–74.
8. C. L. Cardelús, R. K. Colwell, J. E. Watkins, *J. Ecol.* **94**, 144 (2006).
9. G. Brehm, R. K. Colwell, J. Kluge, *Glob. Ecol. Biogeogr.* **16**, 205 (2007).
10. A. R. Wallace, *Tropical Nature and Other Essays* (Macmillan, London, 1878).
11. T. Dobzhansky, *Am. Sci.* **38**, 209 (1950).
12. L. Hannah, R. A. Betts, H. H. Shugart, in *Tropical Rainforest Responses to Climatic Change*, M. B. Bush, J. R. Flenley, Eds. (Springer Praxis, 2007), pp. 351–366.
13. J. R. Flenley, *Clim. Change* **39**, 177 (1998).
14. M. B. Bush, M. R. Silman, D. H. Urrego, *Science* **303**, 827 (2004).
15. F. E. Mayle, D. J. Beerling, W. D. Gosting, M. B. Bush, *Philos. Trans. R. Soc. London B Biol. Sci.* **359**, 499 (2004).
16. M. B. Bush, M. Silman, *J. Quat. Sci.* **19**, 677 (2004).
17. D. D. Ackerly *et al.*, *Bioscience* **50**, 979 (2000).
18. J. Hansen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 14288 (2006).
19. C. A. Deutsch *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 6668 (2008).
20. D. A. Clark, S. C. Piper, C. D. Keeling, D. B. Clark, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5852 (2003).
21. K. J. Feeley, S. Joseph Wright, M. N. Nur Supardi, A. R. Kassim, S. J. Davies, *Ecol. Lett.* **10**, 461 (2007).
22. O. L. Phillips, S. L. Lewis, T. R. Baker, Y. Malhi, in *Tropical Rainforest Responses to Climatic Change*, M. B. Bush, J. R. Flenley, Eds. (Springer Praxis, 2007), pp. 317–332.
23. L. Miles, A. Grainger, O. Phillips, *Glob. Ecol. Biogeogr.* **13**, 553 (2004).
24. S. A. Cowling *et al.*, *Philos. Trans. R. Soc. London B Biol. Sci.* **359**, 539 (2004).
25. J. W. Williams, S. T. Jackson, J. E. Kutzbach, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5738 (2007).
26. M. B. Bush, M. R. Silman, C. McMichael, S. Saatchi, *Philos. Trans. R. Soc. London B Biol. Sci.* **363**, 1795 (2008).
27. R. L. Peters, J. D. S. Darling, *Bioscience* **35**, 707 (1985).
28. G.-R. Walther *et al.*, *Nature* **416**, 389 (2002).
29. C. Ghalambor, R. Huey, P. Martin, J. Tewksbury, G. Wang, *Integr. Comp. Biol.* **46**, 5 (2006).
30. R. G. Pearson, T. P. Dawson, *Glob. Ecol. Biogeogr.* **12**, 361 (2003).
31. S. Williams, E. E. Bolitho, S. Fox, *Proc. R. Soc. London B. Biol. Sci.* **270**, 1887 (2003).
32. J. H. Christensen *et al.*, in *Climate Change 2007: Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon *et al.*, Eds. (Cambridge Univ. Press, Cambridge, 2007), pp. 847–940.
33. J. Kluge, M. Kessler, R. R. Dunn, *Glob. Ecol. Biogeogr.* **15**, 358 (2006).
34. Supported by the Organization for Tropical Studies; the Tropical Ecology Assessment and Monitoring (TEAM) project of Conservation International; University of Connecticut (R.K.C. and C.L.C.); UCLA (A.C.G.); U.S. NSF (DEB-0072702: R.K.C., G.B., and J.T.L.; DEB-0640015: J.T.L.; DEB-0639979: R.K.C.: Research Fellowship and Dissertation Improvement Grant: C.L.C.); Sigma Phi, Explorer's Club, and Steven Vavra Plant Systematics Fund (A.C.G.); and the Deutsche Forschungsgemeinschaft (BR 2280/1-1: G.B.). We thank M. B. Bush, R. L. Chazdon, D. A. Clark, R. R. Dunn, K. M. Kuhn, C. Rahbek, T. F. L. V. B. Rangel, M. R. Silman, and four peer reviewers for comments.

Impact of a Century of Climate Change on Small-Mammal Communities in Yosemite National Park, USA

Craig Moritz,[1,2]* James L. Patton,[1,2] Chris J. Conroy,[1] Juan L. Parra,[1,2] Gary C. White,[3] Steven R. Beissinger[1,4]

We provide a century-scale view of small-mammal responses to global warming, without confounding effects of land-use change, by repeating Grinnell's early–20th century survey across a 3000-meter-elevation gradient that spans Yosemite National Park, California, USA. Using occupancy modeling to control for variation in detectability, we show substantial (~500 meters on average) upward changes in elevational limits for half of 28 species monitored, consistent with the observed ~3°C increase in minimum temperatures. Formerly low-elevation species expanded their ranges and high-elevation species contracted theirs, leading to changed community composition at mid- and high elevations. Elevational replacement among congeners changed because species' responses were idiosyncratic. Though some high-elevation species are threatened, protection of elevation gradients allows other species to respond to migration.

Although human-driven global warming (*1*) has changed phenology of species and contributed to range expansions (*2–6*), contractions of species' ranges are less well documented (*7–10*). Models of future climate-change scenarios predict large range shifts, high global extinction rates, and reorganized communities (*11*, *12*), but model outcomes are also highly uncertain (*13*, *14*). Most studies of species' responses span only a few decades—typically from the 1960 or 1970s, which was a relatively cool period, to the present. Such results can be confounded by decadal-scale climate oscillations (*15*) and landscape modification (*8*, *16*). Furthermore, range shifts are inconsistent when confounded by false absences due to limited historic sampling and inability to control for changes in detectability between sampling periods (*17*, *18*).

We quantified the impact of nearly a century of climate change on the small-mammal community of Yosemite National Park (YNP) in California, USA, by resampling a broad elevational transect (60 to 3300 m above sea level) that Joseph Grinnell and colleagues surveyed from 1914 to 1920 (*19*) (Fig. 1). Their work documented the diversity and distribution of terrestrial vertebrates in California to establish a benchmark for future comparison (*20*), and led to the concept of the ecological niche, the importance of temperature as determinant of range boundaries, and the notion that species respond uniquely to environmental changes (*21*). In contrast to most early–20th century records, the "Yosemite Transect" was densely sampled across elevations (Fig. 1) and is amply documented by specimens (*n* = 4354), field notes (>3000 pages), and photographs (~700) (*22*), enabling precise identification of both species and sampling sites. From daily trapping records, we estimated detectability of species in historical as well as current surveys, permitting the unbiased estimation of species' "absences" from elevational bands in both periods (*23*). The transect spans YNP, a protected landscape since 1890, and allowed us to examine long-term responses to climate change without confounding effects of land-use change, although at low to mid-elevations there has been localized vegetation change relating to seral dynamics, climate change, or both (*24*). Finally, analyses of regional weather records pointed to substantial increase of the average minimum monthly temperature of 3.7°C over the past 100 years, with notable increases from 1910 to 1945 and from 1970 to the present (*15*, *22*) (fig. S1).

[1]Museum of Vertebrate Zoology, University of California, Berkeley, CA 94720, USA. [2]Department of Integrative Biology, University of California, Berkeley, CA 94720, USA. [3]Department of Fish, Wildlife, and Conservation Biology, Colorado State University, Fort Collins, CO 80523, USA. [4]Department of Environmental Science, Policy and Management, University of California, Berkeley, CA 94720, USA.

*To whom correspondence should be addressed. E-mail: craigm@berkeley.edu

Future warming is predicted to cause substantial turnover of species within North American National Parks, including Yosemite (25). Given marked regional warming over the past century, we predicted that species ranges should have shifted upward (5, 10). This should manifest as upward contraction of the lower limit for mid- to high-elevation species, upward shift of the entire range or expansion of the upper limit for low- to mid-elevation species, and altered community composition within elevational bands (9).

Elevational ranges of species and their habitats differed markedly between the gradual western and steep eastern slopes of the transect (19) (Fig. 1). On the west slope, we trapped small mammals at 121 sites compared to 56 in Grinnell's time (table S1), but overall effort and elevational range (~50 to 3300 m) were comparable (22). There were fewer sites on the east side in both time periods (9 for Grinnell, 12 for resurveys) because of limited extent (Fig. 1). Our analyses of richness and turnover focused on species detectable by standardized trapping (37 species) or by observation (6 species; table S2). To test for elevational shifts, we applied occupancy modeling (22, 23) to the 23 west slope taxa with sufficient trapping records to estimate detectability in both periods (tables S1 and S2 and Fig. 2). The best detection model in a set of 36 (table S3) was used to calculate the probability of a false absence (P_{fa}) across trapping sites, where a species was not observed in one sampling period but was in the other (Table 1). Range shifts were significant if $P_{fa} \leq 0.05$. For each species we evaluated eight hypothesized relationships of occupancy, era, and elevation (fig. S2) using the 14 best detection models (table S3) to create model-averaged occupancy-elevation profiles (Fig. 2 and fig. S3). Conservatively, we excluded shifts that were statistically significant but biologically trivial (Fig. 3). In most cases where the P_{fa} test indicated an elevation shift, occupancy models agreed (Table 1 and fig. S3). Exceptions occurred when occupancy models were weak (i.e., insufficient data) or detected changes in occupancy at elevations other than range limits, or when nonstandard data (i.e., records from ad hoc collecting) were included in P_{fa} tests but not in occupancy models.

Elevation limits shifted mostly upward (Table 1 and Fig. 3A), and this occurred more frequently for lower than upper limits ($\chi^2 = 4.26$, df = 1, $P = 0.039$). Twelve of 28 (43%) west slope species showed significant shifts in lower limits, of which 10 increased (mean = +475 m) and two, both shrews, decreased (mean = −744 m). In contrast, upper limits changed significantly in only seven instances, with similar numbers of upward ($n = 4$, mean = +501 m) and downward shifts ($n = 3$, mean = −309 m).
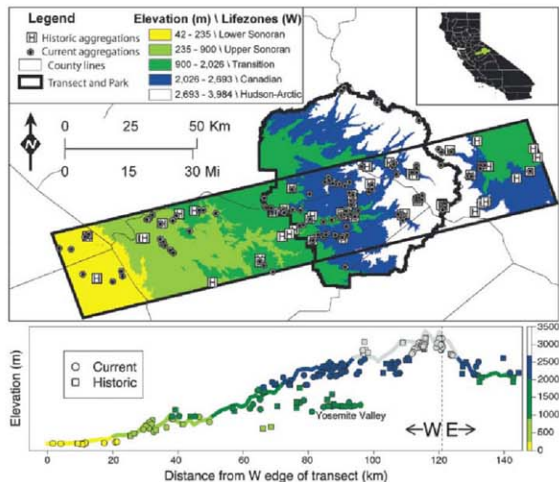
High-elevation species typically experienced range contractions, whereas low-elevation species expanded their ranges upward ($\chi^2 = 8.8$, df = 2, $P = 0.012$), a pattern expected with increased temperature. Lower range limits contracted in 50% of the high-elevation species but in only

10% of low-elevation species, whereas 50% of low-elevation species expanded their upper range compared to none of the high-elevation species (Fig. 3B). High-elevation species contracting (Table 1 and Fig. 2A) included the alpine chipmunk (Tamias alpinus), Belding's ground squirrel (Spermophilus beldingi), water shrew (Sorex palustris), and pika (Ochotona princeps). Range collapse—increased lower limits and decreased upper limits—was observed in two high-elevation species: the bushy-tailed woodrat (Neotoma cinerea) and the shadow chipmunk (T. senex) (Fig. 2B). Parallel trends were observed on the east slope of the Sierra for N. cinerea and S. beldingi (fig. S3). Range contractions due to increases in lower-elevation limits were also observed for two species formerly at mid- to high elevations [the golden-mantled ground squirrel (Spermophilus lateralis) and the long-tailed vole (Microtus longicaudus)] (Table 1). Only one lowland species contracted—the kangaroo rat (Dipodomys heermanni) showed a modest increase in lower limit and a larger decrease in upper limit since Grinnell's time. Range expansions from either expanded upper limits [the pocket mouse (Chaetodipus californicus), the California vole (M. californicus), and the harvest mouse (Reithrodontomys megalotis)] or expanded lower limits (two shrews: Sorex monticola and S. ornatus). Finally, the pinyon mouse (Peromyscus truei) translocated upward (Fig. 2C); both upper and lower limits increased by

~500 m, but it now also occupies montane conifer habitats on the west slope 800 to 1400 m higher after its east slope population expanded upward by ~1000 m to cross the Sierra crest.

Elevational range shifts resulted in modest changes in species richness and composition at varying spatial scales. Species richness averaged across five estimators (26) that account for non-observed species (Fig. 3C, fig. S4, and table S4) declined from the Grinnell era to the present (repeated measures analysis of variance, $F = 32.7$ df = 1, $P = 0.004$). Richness estimators suggest a slight decrease across the whole transect (current-historic mean estimates = −4.4 species, −9%), but not within YNP (+1.3 species, 4%). Species richness was reduced within each life zone, with the largest change in the Lower and Upper Sonoran zones west of YNP. Community similarity between Grinnell's period and the present was high (mean similarity, $S > 0.9$) for the whole transect, the park alone, and most life zones. Species composition was least similar for the Transition and Hudsonian-Arctic zones, as expected given the upward expansions of formerly Sonoran zone taxa and the range shifts of high-elevation species (Table 1).

Closely related species responded idiosyncratically to climate change (Table 1), but why species vary in response is not clear. For example, some species of Peromyscus mice showed elevation range shifts (P. truei), whereas others did not (P. boylii, P. maniculatus). The same is

Legend
- ⊞ Historic aggregations
- ● Current aggregations
- County lines
- ▭ Transect and Park

Elevation (m) \ Lifezones (W)
- 42 – 235 \ Lower Sonoran
- 235 – 900 \ Upper Sonoran
- 900 – 2,026 \ Transition
- 2,026 – 2,693 \ Canadian
- 2,693 – 3,984 \ Hudson-Arctic

Fig. 1. Map of surveyed sites in Grinnell (Historic) and Current surveys relative to the Yosemite National Park boundary and life zones (upper panel), and to an averaged elevational profile (lower panel).

true for chipmunks (*Tamias*), ground squirrels (*Spermophilus*), voles (*Microtus*), and shrews (*Sorex*). Beyond original elevation range (high versus low), life history and ecological traits were weak predictors of which species exhibited upward shifts of their range limits (tables S5 and S6). This was especially true for high-elevation species with upward contraction of their lower range limit. However, lowland species that are short-lived and lay more litters per year (so-called fast life-style species) were more likely to expand their range upward than were their long-lived, less fecund counterparts (table S5 and fig. S5). The elevational replacements among congeners, documented so carefully in the early 20th century (*19*), are now quite different.

By applying occupancy modeling to a thoroughly documented historical record and the re-

survey, we provide an unbiased comparison of changes in species' ranges at the centennial scale. Because much of the transect spans a long-protected National Park, confounding effects of land-use change are minimized. Even so, vegetation has changed within YNP over this period, in part due to fire suppression (*22*). The park was hardly pristine in the early 20th century, with ranching of introduced herbivores in Yosemite Valley and the high country recovering from historical overgrazing. As examples, expansion by *C. californicus* and west slope *P. truei* are associated with fire-related conversion of conifer to shrub habitats, whereas the downward shift of *S. monticola* could reflect recovery of their preferred wet meadow habitats. Increased prevalence of mesic small mammals following cessation of grazing has also been reported for an

analogous community in the Rocky Mountains (*27*).

The preponderance of upward range shifts, leading to contraction of high-elevation species and expansions of low-elevation taxa, accords with the predicted impacts of climate warming (*5, 8, 9*). Although vegetation dynamics have likely contributed to changes at low to mid-elevation, habitat change at higher elevations is limited (*15*) (fig. S6). The ~500-m average increase in elevation for affected species is also consistent with estimated warming of +3°C, assuming a change of temperature with elevation of ~6°C per km. Several small-mammal taxa that responded to changing temperature also showed large range fluctuations during late Quaternary climate fluctuations (*28*), and some have declined regionally (*29*).

Table 1. Analyses of elevation change for 28 west slope species. Given are average detectability per site for Grinnell [*P(G)*] and current [*P(C)*] periods, original elevation range, changes in upper (U) and lower (L) range limit that are significant by the P_{fa} tests, the best supported form of the occupancy model (Elev, elevation; NA, not analyzed), the cumulative Akaike's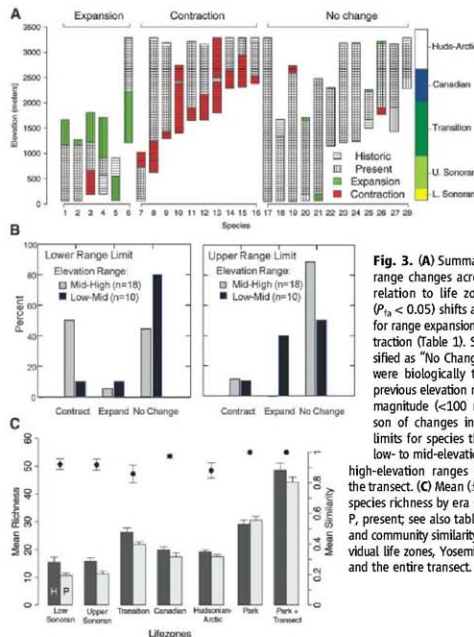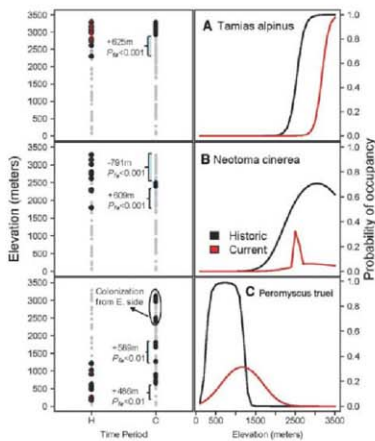 Information Criterion weight for all models with those terms (*w*), and original Lifezone classification (*18*), where L and H refer, respectively, to species with mostly low- to mid-elevation ranges (<2000 m) and mid- to high-elevation ranges (>2000 m) in Grinnell's time; *P. maniculatus* covered the entire transect. Values in bold are further supported by occupancy models. See fig. S4 for elevation plots and models of individual species.

No.	Species	P(G)	P(C)	Original elevation range (m)	Range limit change (m)	Best occupancy model	w	Original life zone (H, L)
				Range expansions				
1	*Microtus californicus*	0.81	0.58	57–1160	+505 U	Elev	0.36	Lower–Upper Sonoran (L)
2	*Reithrodontomys megalotis*	0.99	0.87	57–1160	+112 U	Elev	0.50	Lower–Upper Sonoran (L)
3	*Peromyscus truei**	0.99	0.93	183–1220	**+589 U, +468 L**	Era*(Elev + Elev²)	0.99	Upper Sonoran (L)
4	*Chaetodipus californicus*	0.28	0.19	193–914	**+800 U**	Era*(Elev + Elev²)	0.32	Upper Sonoran (L)
5	*Sorex ornatus*	0.32	0.93	549–914	**−485 L**	Era *(Elev + Elev²)	0.74	Upper Sonoran (L)
6	*Sorex monticolus*	0.99	0.97	2212–3287	**−1003 L**	Era + Elev + Elev²	0.37	Canadian–Hudsonian (H)
				Range contractions				
7	*Dipodomys heermanni*	0.16	0.98	57–1025	+63 L, **−293 U**	Era*Elev	0.48	Lower–Upper Sonoran (L)
8	*Microtus longicaudus*	0.99	0.98	623–3287	**+614 L**	Era + Elev + Elev²	0.74	Transition–Hudsonian (H)
9	*Zapus princeps*	0.98	0.90	1291–3185	+159 L, −64 U	Era + Elev + Elev²	0.53	Transition–Hudsonian (H)
10	*Tamias senex*	0.95	0.71	1402–2743	+1007 L, −334 U	Elev +Elev²	0.48	Canadian (H)
11	*Spermophilus lateralis*	0.70	0.85	1646–3200	**+244 L**	Era*(Elev + Elev²)	0.78	Transition–Hudsonian (H)
12	*Sorex palustris*	0.39	0.23	1658–3155	**+512 L**	Era + Elev + Elev²	0.39	Canadian–Hudsonian (H)
13	*Neotoma cinerea**	0.90	0.71	1798–3287	**+609 L, −719 U**	Era*(Elev + Elev²)	0.83	Canadian–Arctic-Alpine (H)
14	*Spermophilus beldingi**	0.98	0.98	2286–3287	+355 L	Elev	0.32	Canadian–Arctic-Alpine (H)
15	*Tamias alpinus*	0.92	0.95	2307–3353	**+629 L**	Era + Elev	0.56	Hudsonian–Arctic-Alpine (H)
16	*Ochotona princeps*[†]	NA	NA	2377–3871	+153 L	NA	NA	Canadian–Arctic-Alpine (H)
				No change				
17	*Peromyscus maniculatus**	0.99	0.99	57–3287	No change	Era*(Elev + Elev²)	0.72	Lower Sonoran–Arctic-Alpine (H)
18	*Thomomys bottae*[†]	NA	NA	57–1676	No change	NA	NA	Lower Sonoran–Transition (L)
19	*Spermophilus beecheyi*	0.50	0.82	61–2734	−250 U	Era*(Elev + Elev²)	0.89	Lower Sonoran–Canadian (L)
20	*Neotoma macrotis*	0.90	0.91	183–1646	+67 U	Elev + Elev²	0.62	Lower Sonoran–Transition (L)
21	*Peromyscus boylii*	0.98	0.97	183–2469	−122 L	Elev + Elev²	0.60	Upper Sonoran–Transition (L)
22	*Sorex trowbridgii*	0.71	0.88	1160–2286	No change	Elev + Elev²	0.40	Transition–Canadian (H)
23	*Microtus montanus**	0.81	0.98	1217–3155	No change	Elev + Elev²	0.36	Transition–Hudsonian (H)
24	*Tamiasciurus douglasi**[†]	NA	NA	1229–3185	No change	NA	NA	Transition–Hudsonian (H)
25	*Tamias quadrimaculatus*	0.95	0.85	1494–2210	+50 U	Era*(Elev + Elev²)	0.78	Transition–Canadian (H)
26	*Tamias speciosus**	1.00	1.00	1768–3155	+128 L, +65 U	Era*(Elev + Elev²)	1.00	Canadian–Hudsonian (H)
27	*Thomomys monticola*[†]	NA	NA	1905–3155	No change	NA	NA	Canadian–Hudsonian (H)
28	*Marmota flaviventris*[†]	NA	NA	2469–3353	No change	NA	NA	Canadian–Arctic-Alpine (H)

* Similar trends are observed for east-side populations (see fig. S4). †These species were encountered by observation and/or specialized trapping and were not subject to occupancy analyses.

Fig. 2. Example elevation plots from the west slope transect of upward range expansion (*T. alpinus* and *P. truei*) (**A** and **C**), and range collapse (*N. cinerea*) (**B**). Shown are occupied (black) and unoccupied (gray) sites, probability of false absence (P_{fa}), and model-averaged occupancy-elevation profiles (table S3 and fig. S2). *P. truei* colonized high elevations west of the Sierra crest from the eastern slope. Red marks for historical elevation profile of *T. alpinus* refer to ad hoc records.

Recent trends do not bode well for several mid- to high-elevation species, including some endemic to the high Sierra (e.g., *T. alpinus*) (Fig. 3A). Nevertheless, species diversity within Yosemite has changed little, because range expansions compensated for retractions. Our results confirm that protecting large-scale elevation gradients retains diversity by allowing species to migrate in response to climate and vegetation change. The long-recognized importance of protected landscapes has never been greater.

References and Notes

1. Intergovernmental Panel on Climate Change, *Climate Change 2007: Synthesis Report Contribution of Working Groups I, II, and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (IPCC, Geneva, Switzerland, 2007).
2. C. Rosenzweig et al., *Nature* 453, 353 (2008).
3. C. Parmesan, G. Yohe, *Nature* 421, 37 (2003).
4. G. R. Walther et al., *Nature* 416, 389 (2002).
5. C. Parmesan, *Annu. Rev. Ecol. Evol. Syst.* 37, 637 (2006).
6. J. Lenoir, J. C. Gégout, P. A. Marquet, P. de Ruffray, H. Brisse, *Science* 320, 1768 (2008).
7. R. J. Rowe, *J. Biogeogr.* 32, 1883 (2005).
8. R. Hickling, D. B. Roy, J. K. Hill, R. Fox, C. D. Thomas, *Glob. Change Biol.* 12, 450 (2006).
9. R. J. Wilson, D. Gutiérrez, J. Gutiérrez, V. J. Monserrat, *Glob. Change Biol.* 13, 1873 (2007).
10. C. D. Thomas, A. M. A. Franco, J. K. Hill, *Evolution Int. J. Org. Evolution* 21, 415 (2006).
11. C. D. Thomas et al., *Nature* 427, 145 (2004).
12. J. W. Williams, S. T. Jackson, *Front. Ecol. Environ* 5, 475 (2007).
13. A. J. Davis, L. S. Jenkinson, J. H. Lawton, B. Shorrocks, S. Wood, *Nature* 391, 783 (1998).
14. R. G. Pearson et al., *J. Biogeogr.* 33, 1704 (2006).
15. C. I. Millar, R. D. Westfall, D. L. Delany, J. C. King, L. J. Graumlich, *Arct. Antarct. Alp. Res.* 36, 181 (2004).
16. W. Jetz, D. S. Wilcove, A. P. Dobson, *PLoS Biol.* 5, e157 (2007).
17. J. K. Hill et al., *Proc. R. Soc. Lond. B. Biol. Sci.* 269, 2163 (2002).
18. L. P. Shoo, S. E. Williams, J. M. Hero, *Biol. Conserv.* 125, 335 (2005).
19. J. Grinnell, T. Storer, *Animal Life in the Yosemite* (Univ. of California Press, Berkeley, CA, 1924).
20. J. Grinnell, *Pop. Sci. Mon.* 77, 163 (1910).
21. J. Grinnell, *Am. Nat.* 51, 115 (1917).
22. See supporting material on Science Online.
23. D. I. MacKenzie, *Inferring Patterns and Dynamics of Species Occurrence* (Elsevier, San Diego, 2006).
24. A. Guarin, A. H. Taylor, *For. Ecol. Manage.* 218, 229 (2005).
25. C. E. Burns, K. M. Johnston, O. J. Schmitz, *Proc. Natl. Acad. Sci. U.S.A.* 100, 11474 (2003).
26. A. Chao, in *Encyclopedia of Statistical Sciences*, N. Balakrishnan, C. B. Read, B. Vidakovic, Eds. (Wiley, New York, 2005).
27. R. J. Rowe, *Am. Nat.* 170, 242 (2007).
28. D. K. Grayson, *Quat. Sci. Rev.* 25, 2964 (2006).
29. E. A. Beever, P. E. Brussard, J. Berger, *J. Mammal.* 84, 37 (2003).
30. Supported by funding from the Yosemite Foundation, the National Parks Service, and NSF (DEB 0640859). Assistance was received from A. Chang, H. Shafi, K. Tsao, D. Yang, W. Monahan, R. Hijmans, and M. Koo. L. Chow (U.S. Geological Survey Yosemite) and S. Thompson (Yosemite National Park) stimulated and participated in the project. Comments from R. Colwell, M. Power, F. Hamer, and L. Shoo improved this manuscript.

Fig. 3. (**A**) Summary of elevational range changes across all species in relation to life zones. Significant ($P_{fa} < 0.05$) shifts are colored green for range expansion and red for contraction (Table 1). Species were classified as "No Change" if range shifts were biologically trivial (<10% of previous elevation range) or of small magnitude (<100 m). (**B**) Comparison of changes in elevation-range limits for species that formerly had low- to mid-elevation versus mid- to high-elevation ranges (Table 1) across the transect. (**C**) Mean (± SE) estimates of species richness by era (bars: H, historic; P, present; see also table S4 and fig. S4) and community similarity (points) for individual life zones, Yosemite National Park, and the entire transect.

Small Molecule–Induced Allosteric Activation of the *Vibrio cholerae* RTX Cysteine Protease Domain

Patrick J. Lupardus,[1]* Aimee Shen,[3]* Matthew Bogyo,[3]† K. Christopher Garcia[1,2]†

Vibrio cholerae RTX (repeats in toxin) is an actin-disrupting toxin that is autoprocessed by an internal cysteine protease domain (CPD). The RTX CPD is efficiently activated by the eukaryote-specific small molecule inositol hexakisphosphate (InsP$_6$), and we present the 2.1 angstrom structure of the RTX CPD in complex with InsP$_6$. InsP$_6$ binds to a conserved basic cleft that is distant from the protease active site. Biochemical and kinetic analyses of CPD mutants indicate that InsP$_6$ binding induces an allosteric switch that leads to the autoprocessing and intracellular release of toxin-effector domains.

Most secreted bacterial toxins are produced as inactive precursors that become proteolytically activated upon entering a eukaryotic cell (*1*). A select group of these toxins undergo autoproteolysis upon entry into the host cytosol, resulting in the release of their effector domains (*2*, *3*). The *Vibrio cholerae* RTX (repeats in toxin) is a member of the multifunctional autoprocessing RTX (MARTX) family of toxins, which all contain a cysteine protease domain (CPD) predicted to mediate the autoproteolytic activation of the secreted protoxin upon entry into the eukaryotic cytosol (*4*). Almost all clinical and environmental isolates of *V. cholerae* produce RTX (*5*), which enhances virulence and
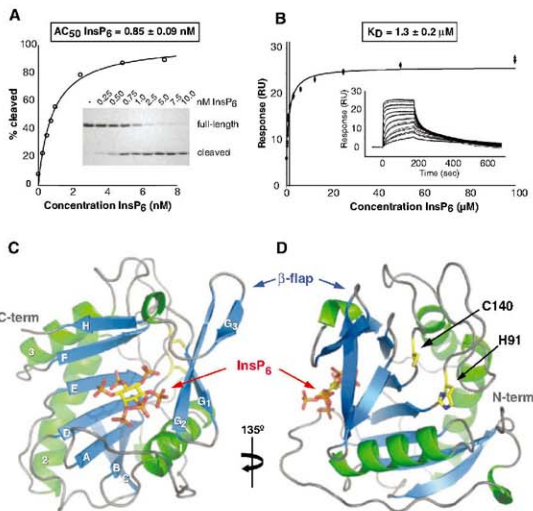
colonization in murine models of *V. cholerae* infection (*6*, *7*). RTX autoprocessing is thought to release its actin-disrupting effector domains from the target cell's plasma membrane into the cytosol (*4*) (fig. S1). Although autoproteolysis is essential for RTX toxin function (*3*), the mechanism of RTX CPD activation is unclear.

It has recently been shown that the small molecules guanosine 5′-triphosphate (GTP) and inositol hexakisphosphate (InsP$_6$) stimulate autoprocessing of the RTX CPD (*3*, *8*); also, InsP$_6$ activates a distantly related protease domain in toxin B of *Clostridium difficile* (*2*, *9*). We tested the ability of guanosine 5′-O-(3′-thiotriphosphate) (GTP-γ-S), InsP$_6$, and InsP$_6$ metabolites to ac-

tivate *V. cholerae* RTX CPD autocleavage in vitro. InsP$_6$ potently activated autocatalysis, with a half-maximal autocleavage concentration (AC$_{50}$) of 0.9 nM versus 0.19 μM, 0.72 μM, and 240 μM for InsP(1,4,5,6)$_4$, InsP(1,3,4,5,6)$_5$, and GTP-γ-S, respectively (Fig. 1A and fig. S3). We confirmed the interaction between InsP$_6$ and RTX CPD using surface plasmon resonance (SPR), which indicated that InsP$_6$ binds to the RTX CPD with an equilibrium binding affinity constant (K_d) of 1.3 ± 0.2 μM (Fig. 1B).

To gain structural insight into the InsP$_6$-mediated activation of RTX CPD, we purified an autocleaved minimal RTX CPD catalytic domain for cocrystallization with InsP$_6$ (*10*). We determined the structure of the RTX CPD at a resolution of 2.1 Å, consisting of amino acids 5 through 203 (with residue 1 being the P1′ alanine and P1′ referring to the residue C-terminal to the scissile bond). The protease domain comprises a seven-stranded beta sheet, with five central parallel strands and two antiparallel capping strands (Fig. 1, C and D). Two helices flank one

[1]Department of Molecular and Cellular Physiology and Department of Structural Biology, Stanford University School of Medicine, Stanford, CA 94305, USA. [2]Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305, USA. [3]Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305, USA.

*These authors contributed equally to this work.
†To whom correspondence should be addressed. E-mail: mbogyo@stanford.edu (M.B.); kcgarcia@stanford.edu (K.C.G.)

Fig. 1. InsP$_6$ activates the *V. cholerae* RTX CPD and the architecture of the InsP$_6$-CPD complex. (**A**) Activation of RTX cysteine protease domain autocleavage by InsP$_6$. Recombinant RTX CPD (amino acids 3391 to 3650) was incubated with the indicated concentrations of InsP$_6$ for 2 hours, and autocleavage was assessed by SDS–polyacrylamide gel electrophoresis (SDS-PAGE) and Coomassie staining (representative gel inset). Reactions were performed in triplicate, and the amount of autocleaved protein relative to the total protein amount was analyzed by densitometry, averaged and plotted versus concentration of InsP$_6$. 50% of the wild-type CPD was autoprocessed (AC$_{50}$) at 0.85 ± 0.02 nM InsP$_6$ (mean ± SD). (**B**) SPR analysis of InsP$_6$ binding to wild-type biotinylated CPD immobilized on a streptavidin-coupled surface. Representative sensorgram (inset) shows the binding of InsP$_6$ to the CPD-bound surface over a concentration of 0.1 to 100 μM. Equilibrium binding analysis indicates a K_d of 1.3 ± 0.2 μM (SD). (**C**) Structure of the CPD-InsP$_6$ complex viewed from above the InsP$_6$-binding site. (**D**) A view of the structure rotated ~135° to show the active site. The InsP$_6$-binding and active sites are separated by a 22–amino acid β-hairpin structure (β-flap). InsP$_6$ and the side chains of the catalytic dyad (Cys140/His91) are shown as stick models.

Fig. 2. The InsP$_6$-binding and active sites. (**A**) Electrostatic surface potential of the CPD as viewed from above the InsP$_6$-binding site. Blue denotes a positively charged surface; red denotes a negatively charged surface. InsP$_6$ is shown in the binding site as a stick model. (**B**) Close-up view of the InsP$_6$-binding site. Side chains that directly interact with InsP$_6$ are labeled and shown as yellow sticks. The electron density for InsP$_6$ ($2F_{obs} - F_{calc}$) is contoured at 2σ. (**C**) Surface topology of the CPD active site. The P1 substrate pocket, C140, and H91 are highlighted in orange, yellow, and blue, respectively. The N terminus is shown as a yellow ribbon, terminating at Ile5 and highlighting the threading of this region along the surface of the core domain. The remaining residues not visible at the N terminus are depicted as a yellow dashed line to illustrate the approximate positioning of the chain during catalysis. (**D**) Close-up view of the P1 substrate pocket. Amino acids that line the pocket are labeled and colored orange. InsP$_6$ is shown as in (B) to demonstrate the position of the catalytic site with respect to the InsP$_6$-binding site.
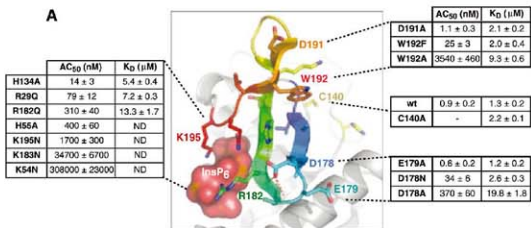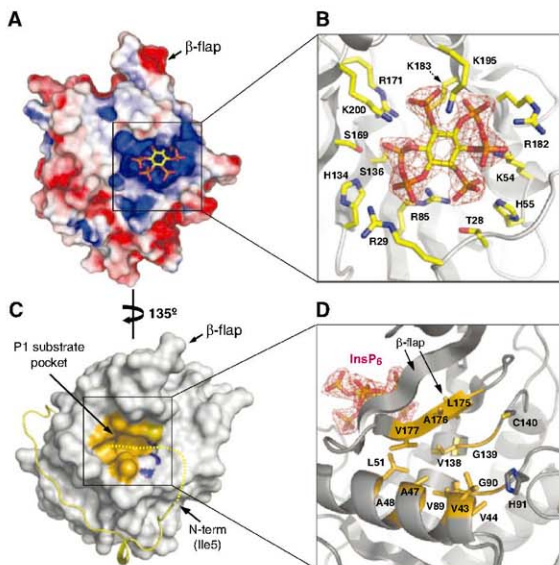
Fig. 3. β-Flap mutations decouple CPD autocatalysis and RTX activity from InsP$_6$ binding. (**A**) Comparison of autocleavage efficiency (AC$_{50}$) versus InsP$_6$ binding (K_d) measured by SPR for mutations in the InsP$_6$-binding site (left table) and β-flap (right tables, top and bottom). The β-flap region of the CPD is rainbow-colored, starting with blue at the N-terminal end. The β-flap, catalytic site, and visible InsP$_6$-interacting side chains are shown as sticks. Data are expressed as mean ± SD. ND, not determinable. (**B**) Western blot analysis of RTX in supernatant harvested from log-phase *V. cholerae* cultures. Supernatants from *V. cholerae* strains harboring either an intact *rtxA* gene (wt), a null mutation in *rtxA* (Δ*rtxA*), or point mutations in the region encoding the CPD domain of RTX (*C140A* is catalytic-dead; *R182Q/K183N* is mutated at two InsP$_6$-binding residues; and *W192A* is a β-flap mutation) were blotted using an anti-CPD antibody. (**C**) Actin crosslinking induced upon incubation of RTX with HFF cells. *V. cholerae* strains used in (A) were incubated with HFFs for 90 min, then the HFF cells were lysed. Actin crosslinking was visualized by SDS-PAGE and Western blotting by using an actin-specific antibody. The crosslinked forms of actin are labeled to the right.
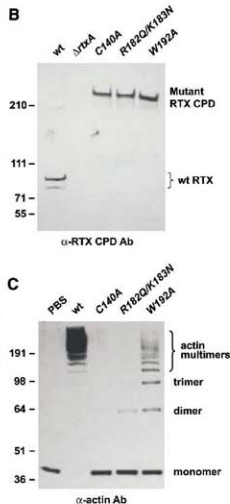
side of the sheet, lying diagonally in a groove created by a ~90° twist of the sheet, and a third helix caps the other side of the sheet. The Cys-His catalytic dyad (C140 and H91) lies at the C-terminal ends of the central D and E β strands. The overall structure is reminiscent of the clan CD family of cysteine proteases, which suggests a common ancestor for this family of enzymes. Comparison with the two most closely related known protease structures, human caspase-7 (11) and *Porphyromonas gingivalis* gingipain-R (12) (fig. S4), reveals that two C-terminal helices have been replaced by a three-strand flap. This "β-flap" structure forms a cleft in which we identified electron density for a single InsP$_6$ molecule (fig. S5).

The InsP$_6$-binding pocket is lined with basic residues, burying approximately 890 Å^2 of surface area (Fig. 2A). Nine of the twelve residues that directly interact with InsP$_6$ via hydrogen bond contacts are positively charged, with a core of six residues (K54, R85, S136, R171, K183, and K200) forming the bottom of the pocket and five others surrounding the InsP$_6$ molecule (T28, R29, H55, H134, S169, and R182) (Fig. 2B and fig. S6) (13). K195 covers the top of the InsP$_6$ molecule, interacting with the C1, C5, and C6 phosphate groups. The protein–InsP$_6$ interface is further stabilized by a network of water molecules between InsP$_6$ and the β-flap. On the opposite side of the β-flap is the catalytic dyad and a large hydrophobic pocket for the P1 amino acid (the residue N-terminal to the scissile bond) (Fig. 2C). The β-flap contributes three of the twelve hydrophobic amino acids that line one side of the P1 pocket, and helix 1 and two central beta sheets (D and E) contribute the remainder (Fig. 2D). The surface properties and size of the P1 substrate pocket are consistent with the cleavage of RTX CPD after an N-terminal leucine (3). The final N-terminal residue observed in our structure is I5 (the P5' position) (Fig. 2C), which lies ~14 Å from the catalytic cysteine. No electron density was observed for the P1' through P4' positions, suggesting that they do not make strong contacts with the protease and may minimally contribute to substrate specificity.

The InsP$_6$-binding site is structurally segregated from the active site, which clearly indicates that InsP$_6$ does not act as a cofactor for catalysis. Indeed, kinetic analyses revealed that InsP$_6$ binding is independent of substrate binding to the active site and that the concentration of InsP$_6$ does not alter the affinity of the InsP$_6$-bound enzyme for its substrate (fig. S7). Because RTX CPD activity was strictly dependent on InsP$_6$ in these analyses, we hypothesized that InsP$_6$ binding may regulate exposure of the active site. We therefore tested the ability of a fluorescent maleimide derivative to alkylate the catalytic cysteine of wild-type CPD and a mutant CPD lacking two InsP$_6$-interacting residues (R182Q/K183N) (14). Although weak fluorescent labeling was observed for both wild-type and mutant CPD in the absence of InsP$_6$, dose-

dependent labeling of the active-site cysteine was observed in the presence of InsP$_6$ only for the wild-type CPD (fig. S8). Thus, productive binding of InsP$_6$ is required to expose the active-site cysteine of the protease to substrates and inhibitors. Consistent with this observation, pretreatment of wild-type CPD immobilized on an SPR chip with *N*-ethylmaleimide (NEM) alone failed to block InsP$_6$-induced autocleavage of the CPD from the chip (fig. S9). Simultaneous treatment of the CPD with InsP$_6$ and NEM, however, inhibited InsP$_6$-induced autocleavage, which indicated that NEM can only react with the active-site cysteine in the presence of InsP$_6$. These results strongly suggest an allosteric mechanism of activation in which the active site is disordered or occluded in the absence of InsP$_6$, a mode of regulation that likely protects the protease active-site sulfhydryl until the toxin enters a eukaryotic cell.

Inspection of the structure suggested that the β-flap, which lines the side of the InsP$_6$-binding cleft closest to the catalytic site, may contribute to enzyme activation by properly ordering the P1 pocket and active site. Among the many side chains that coordinate InsP$_6$, the β-flap contains three residues (R182, K183, and K195) that form a three-pronged "clamp" above and below InsP$_6$ (Figs. 2A and 3A). We tested the effect of the mutations R182Q, K183N, and K195N on both InsP$_6$ binding (fig. S10) and catalytic activity (fig. S11) relative to wild-type CPD (Fig. 3A). Mutation of K183 and K195 abrogated both InsP$_6$ binding and autocleavage activity (Fig. 3A, left table), whereas mutation of R182 only moderately reduced InsP$_6$ binding but dramatically decreased autocatalysis as compared with the wild type. R182 not only binds InsP$_6$, it also engages in structurally stabilizing hydrogen-bonding interactions with D24 (fig. S12). This may explain the more considerable impact of the R182Q mutation on catalysis rather than InsP$_6$ binding, because R182 may primarily fine-tune the structure of the β-flap and may contribute nominally to binding InsP$_6$. Thus, diminution of InsP$_6$ binding on one side of the flap clearly reduces catalysis mediated by residues on the opposite side of the flap.

We sought to determine if the mutation of non–InsP$_6$-interacting residues in the β-flap might "decouple" InsP$_6$-binding affinity from the autocatalytic activity of the enzyme. We focused on W192 in strand G$_3$, which makes van der Waals contacts with strands G$_1$ and G$_2$ (Fig. 3A), and D178, which stabilizes the G$_1$-G$_2$ β-hairpin by hydrogen bonding with three backbone amide nitrogens and helps anchor the center of the G$_2$ strand by forming a salt bridge with the side chain of H184 (Fig. 3A). We made conservative (W192F and D178N) and more potent (W192A and D178A) mutations; as controls, we also mutated nearby solvent-facing residues D191 and E179 (D191A and E179A). The conservative mutations (W192F and D178N) minimally altered InsP$_6$-binding affinity yet considerably reduced autocatalysis relative to the wild type and controls D191A and E179A (Fig. 3A, right). Fur-

thermore, the more potent W192A and D178A mutations induced only a modest drop in InsP$_6$-binding affinity yet a more severe defect in catalysis. The D178A mutant exhibited InsP$_6$ binding and catalytic activity similar to those of the R182Q mutant, consistent with their similar positions in the β-flap. The W192A mutant was most dramatically affected; despite having moderate affinity for InsP$_6$ (9.3 ± 0.6 μM), W192A had a catalytic defect that was twice as strong as that of the InsP$_6$-binding mutant K195N. Thus, mutation of residues that apparently communicate InsP$_6$ binding to the active site through structural rearrangement of the β-flap can decouple InsP$_6$ binding from enzyme activation. The proposed structural rearrangement, however, appears to be subtle because the circular dichroism (CD) spectra of the free- and InsP$_6$-bound enzymes were nearly superimposable (fig. S13).

To assess the biological importance of InsP$_6$ binding and β-flap integrity on RTX CPD function in vivo, we constructed *V. cholerae* strains containing CPD point mutations C140A (catalytic dead), R182Q/K183N (InsP$_6$-binding defective), and W192A (β-flap–transition defective). Western detection of the CPD (Fig. 3B) showed that although most wild-type RTX was autoprocessed during the growth of *V. cholerae* in LB medium, RTX containing catalytic dead (C140A), InsP$_6$-binding (R182Q/K183N), or β-flap (W192A) mutations were unprocessed (Fig. 3B). Thus, even in the presence of high levels of InsP$_6$ in LB medium, all three mutations prevented CPD activation in the native RTX. When *V. cholerae* mutant strains or supernatants were incubated with human foreskin fibroblast (HFF) cells, intracellular actin cross-linking induced by the mutant strains R182Q/K183N and W192A was severely reduced relative to the wild type, whereas C140A failed to induce actin cross-linking (Fig. 3C and fig. S14). Thus, InsP$_6$ binding and an intact β-flap are required for RTX autocleavage and effector function.

The InsP$_6$-interacting residues in related MARTX cysteine protease domains are almost invariably conserved (fig. S2), which strongly suggests a shared mechanism for toxin activation in Gram-negative bacteria. Because InsP$_6$ is exclusive to eukaryotes (15) and is present at cytosolic concentrations >10 μM (16), the responsiveness of the MARTX family to InsP$_6$ through the evolution of a proteolytic biosensor seems an ingenious strategy for regulating the function of a secreted toxin.

References and Notes
1. V. M. Gordon, S. H. Leppla, *Infect. Immun.* **62**, 333 (1994).
2. J. Reineke *et al.*, *Nature* **446**, 415 (2007).
3. K. L. Sheahan, C. L. Cordero, K. J. Satchell, *EMBO J.* **26**, 2552 (2007).
4. K. J. Satchell, *Infect. Immun.* **75**, 5079 (2007).
5. C. L. Cordero, S. Sozhamannan, K. J. Satchell, *J. Clin. Microbiol.* **45**, 2289 (2007).
6. G. Olivier, G. K. Haines 3rd, Y. Tan, K. J. Satchell, *Infect. Immun.* **75**, 5035 (2007).
7. G. Olivier, N. H. Salzman, K. J. Satchell, *Infect. Immun.* **75**, 5043 (2007).
8. K. Prochazkova, K. J. Satchell, *J. Biol. Chem.* **283**, 23656 (2008).

9. M. Egerer, T. Giesemann, T. Jank, K. J. Satchell, K. Aktories, *J. Biol. Chem.* **282**, 25314 (2007).
10. Materials and methods are available as supporting material on Science Online.
11. J. A. Hardy, J. Lam, J. T. Nguyen, T. O'Brien, J. A. Wells, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 12461 (2004).
12. A. Eichinger et al., *EMBO J.* **18**, 5453 (1999).
13. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
14. In the mutants, other amino acids were substituted at certain locations; for example, R182Q indicates that arginine at position 182 was replaced by glutamine.
15. R. H. Michell, *Nat. Rev. Mol. Cell Biol.* **9**, 151 (2008).
16. R. F. Irvine, M. J. Schell, *Nat. Rev. Mol. Cell Biol.* **2**, 327 (2001).
17. We thank S. Juo for assistance with data collection and structure determination, E. D. Sandoval for assistance with kinetic analyses and helpful discussion, M. Blokech and G. Schoolnik for help with V. cholerae strain construction, and A. Guzzetta for intact mass analysis of SeMet-labeled CPD. P.J.L. is a Damon Runyon Fellow, supported by the Damon Runyon Cancer Research Foundation. K.C.G. is supported by the NIH Keck Foundation and the Howard Hughes Medical Institute. M.B. is supported by the Burroughs Wellcome Foundation, the Searle Scholars Program, and the NIH National Technology Center for Networks and Pathways (grant U54RR020843). P.J.L, A.S., M.B., and K.C.G. are listed as inventors on a patent application related to use of the V. cholerae RTX CPD for biotechnical applications. Coordinates and structure factors have been deposited in the Protein Data Bank (www.rcsb.org) under accession number 3EEB.

Noncytotoxic Lytic Granule–Mediated CD8+ T Cell Inhibition of HSV-1 Reactivation from Neuronal Latency

Jared E. Knickelbein,[1,2,3] Kamal M. Khanna,[3]* Michael B. Yee,[3] Catherine J. Baty,[4,5] Paul R. Kinchington,[3,6] Robert L. Hendricks[3,6,7]†

Reactivation of herpes simplex virus type 1 (HSV-1) from neuronal latency is a common and potentially devastating cause of disease worldwide. CD8+ T cells can completely inhibit HSV reactivation, with interferon-γ affording a portion of this protection. We found that CD8+ T cell lytic granules are also required for the maintenance of neuronal latency both in vivo and in ex vivo ganglia cultures and that their directed release to the junction with neurons in latently infected ganglia did not induce neuronal apoptosis. Here, we describe a nonlethal mechanism of viral inactivation in which the lytic granule component, granzyme B, degrades the HSV-1 immediate early protein, ICP4, which is essential for further viral gene expression.

S everal lines of evidence support a role for CD8+ T cells in controlling herpes simplex virus type 1 (HSV-1) latency. CD8+ T cells, many expressing granzyme B (GrB), are found juxtaposed to HSV-1 latently infected sensory neurons of both humans (1–4) and mice (5–8). In C57BL/6 mice, CD8+ T cells specific for the immunodominant HSV-1 glycoprotein $B_{498-505}$ epitope (gB-CD8) polarize their T cell receptor (TCR) to junctions with neurons in situ forming apparent immunologic synapses (9). Murine gB-CD8 can block HSV-1 reactivation from latency in vivo and in ex vivo ganglia cultures in a major histocompatibility complex–dependent manner (9–11). Because HSV-1 estab-

lishes latency solely within ganglionic neurons (12, 13), we hypothesize that some latently infected neurons directly present viral antigens to HSV-specific CD8+ T cells during attempted reactivation, which is subsequently quelled by CD8+ T cell effector functions.

CD8+ T cells can use interferon-γ (IFN-γ) to block HSV-1 reactivation in some, but not all, latently infected sensory neurons (14, 15). HSV-1 reactivation is suppressed by CD8+ T cells in neurons that are refractory to IFN-γ through an as yet undefined mechanism. Lytic granules represent an important CD8+ T cell effector mechanism, but their use is generally lethal to targeted cells. GrB-expressing gB-CD8 from latently infected trigeminal ganglia (TG) polarized and released their lytic granules toward susceptible fibroblasts, leading to apoptosis (fig. S1) (16). Thus, we investigated whether gB-CD8 used lytic granules during immunosurveillance of latently infected neurons and also whether they induced neuronal apoptosis.

GrB+ gB-CD8 expanded from latently infected TG of wild-type (WT) mice (fig. S2) were added to cultures of dispersed TG in which reactivated HSV-1 had spread to the surrounding fibroblasts. Most fibroblasts targeted by gB-CD8 showed active caspase staining in punctate, multi-focal, or diffuse patterns (fig. S3, A and C), which is consistent with early, intermediate, and late stages of apoptosis, respectively (17). Conversely, none of the gB-CD8–targeted neurons showed

caspase activation (fig. S3, B and C). Thus, either CD8+ T cells do not release lytic granules toward neurons, or lytic granule release does not activate the caspase system of neurons.

To distinguish between these possibilities, we first documented CD8+ T cell polarization of GrB toward junctions with neurons in latently infected TG in situ (Fig. 1A) and ex vivo (Fig. 1B), suggesting ongoing use of directed lytic granule release by CD8+ T cells during immunosurveillance of latently infected ganglia. Histologic studies of HSV-1 latently infected human (1–4) and murine (5–8) ganglia have failed to detect morphological signs of apoptosis in neurons in direct contact with activated CD8+ T cells. To directly investigate whether neurons are refractory to lytic granule–mediated apoptosis, WT GrB-expressing gB-CD8 were added to dispersed latently infected TG directly ex vivo when HSV-1 is confined to neurons. Of 13 documented neuron/CD8+ T cell interactions exhibiting lytic granule release, none of the targeted neurons exhibited activated caspases (Fig. 1C), whereas neuronal caspases could be activated by ethanol treatment (Fig. 1D). CD8+ T cells contacting nonneuronal cells or not contacting any cell showed no lytic granule release (Fig. 1E). The selective resistance of neurons to apoptosis induction by CD8+ T cell lytic granules might be

Table 1. Pfn−/− and GrB−/− gB-CD8 are less efficient than WT gB-CD8 at inhibiting HSV-1 reactivation in a common pool of latently infected neurons. HSV-1 reactivation detected by plaque assay of supernatants from dispersed 34- to 44-dpi WT TG depleted of endogenous CD8+ T cells and cultured at one-fifth TG per well with the indicated number and type of gB-CD8 from culture initiation. +, reactivation detected; −, no reactivation detected; x, condition not tested. Data were pooled from four independent experiments (n = 15 to 35 cultures per condition).

Number of gB-CD8 per well	Type of gB-CD8		
	WT	Pfn−/−	GrB−/−
2×10^3	+	x	x
2×10^4	−	+	+
4×10^4	−	+	+
5×10^4	−	+	−
1×10^5	−	+	−

[1]Graduate Program in Immunology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA. [2]Medical Scientist Training Program, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA. [3]Department of Ophthalmology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA. [4]Center for Biologic Imaging, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA. [5]Department of Cell Biology and Physiology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA. [6]Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA. [7]Department of Immunology, University of Pittsburgh School of Medicine, Pittsburgh, PA 15213, USA.

*Present address: Department of Immunology, University of Connecticut, Farmington, CT 06030, USA.
†To whom correspondence should be addressed. E-mail: hendricksrr@upmc.edu

due to the anti-apoptotic activity attributed to HSV-1 latency-associated transcripts (18).

The use of lytic granules in maintaining HSV-1 latency in vivo was assessed by infecting the corneas of WT mice or mice deficient in the lytic granule components perforin (Pfn$^{-/-}$) or GrB (GrB$^{-/-}$). All three strains cleared virus from infected corneas and TG with similar kinetics (fig. S4) and initially retained a similar latent viral load in the TG (Fig. 2A). However, latency was unstable in Pfn$^{-/-}$ and GrB$^{-/-}$ TG, as indicated by a significant increase in the number of viral genome copies compared with WT TG at 14 dpi (Fig. 2A). The viral load returned to WT levels in GrB$^{-/-}$ mice by 20 days post-infection (dpi) and in Pfn$^{-/-}$ mice by 34 to 36 dpi, suggesting additional mechanisms (such as INF-γ) in blocking HSV-1 reactivation. HSV-1 reactivation frequency correlates directly with viral load and inversely with CD8$^+$ T cell numbers in ex vivo TG cultures (19). At >30 dpi, these parameters became equivalent in WT, Pfn$^{-/-}$, and GrB$^{-/-}$ TG (Fig. 2A and fig. S5), yet cultures

of Pfn$^{-/-}$ and GrB$^{-/-}$ TG reactivated to a significantly greater extent than did those of WT TG (Fig. 2B), further establishing a role for Pfn and GrB in inhibiting HSV-1 reactivation. To assess the relative roles of lytic granules and IFN-γ in blocking HSV-1 reactivation, dispersed latently infected WT or Pfn$^{-/-}$ TG were cultured in medium alone or medium supplemented with recombinant IFN-γ (rIFN-γ) or neutralizing antibodies against IFN-γ or CD8α (Fig. 2C). The extent of reactivation in Pfn$^{-/-}$ TG was further elevated by neutralization of IFN-γ to levels similar to CD8-neutralized WT TG. The addition of rIFN-γ decreased the level of reactivation in Pfn$^{-/-}$ TG, although it remained higher than WT TG containing endogenous CD8$^+$ T cells. Thus, both lytic granules and IFN-γ are important in immune control of HSV-1 reactivation from latency.

Direct evidence that gB-CD8 use lytic granules to inhibit HSV-1 reactivation came from the observation that 2.5-fold more GrB$^{-/-}$ and greater than fivefold more Pfn$^{-/-}$ gB-CD8 were required to block HSV-1 reactivation in cultures of pooled, CD8-depleted

WT TG (Table 1). Thus, gB-CD8 use lytic granules to block HSV-1 reactivation, and GrB is an important (but not the only) lytic granule component involved. An additional contribution from GrA is likely because this granzyme inhibits the interneuronal spread of HSV-1 within infected ganglia (20).

We hypothesized that GrB might inhibit the viral life cycle without killing the host neuron by directly cleaving an essential HSV-1 protein similar to GrH cleavage of the adenovirus DNA-binding protein (21). The GraBCas bioinformatics program (22) identified two distinct GrB cleavage sites in ICP4 (fig. S6), an HSV-1 immediate early (IE) protein that is essential for efficient viral transcription beyond the α genes (23). Levels of ICP4 were reduced in HSV-1–infected syngenic fibroblasts after exposure to WT but not Pfn$^{-/-}$ or GrB$^{-/-}$ gB-CD8 (Fig. 3, A and B). GrB-mediated degradation of ICP4 was directly established by exposing lysates of 293T cells transfected to express ICP4, lysates of fibroblasts infected with HSV-1, or ICP4 immunoprecipitated from
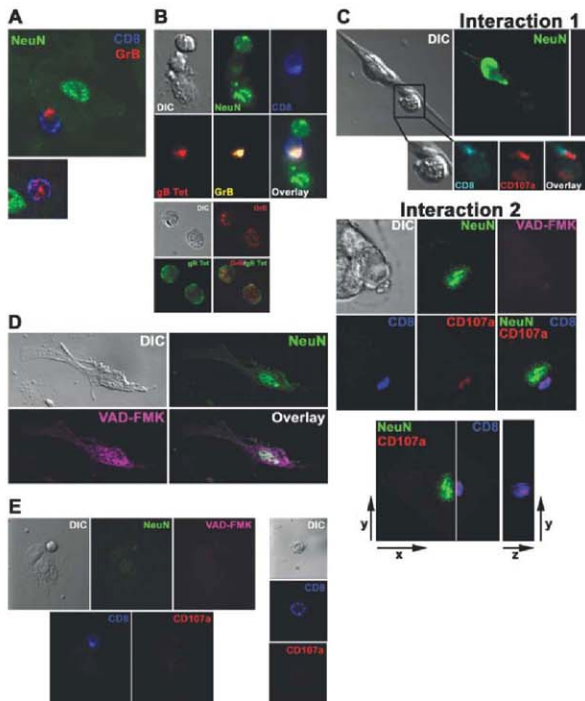
Fig. 1. CD8$^+$ T cells release lytic granules toward neurons within HSV-1 latently infected ganglia without activating neuronal caspases. (**A**) In situ confocal images of intact latently infected WT TG stained with antibodies to CD8α, GrB, and NeuN (neuronal nucleus). (Top) CD8$^+$ T cell with GrB polarized to the junction with a NeuN$^+$ neuron. (Bottom) Most CD8$^+$ T cells have GrB dispersed throughout the cell. (**B**) Imaging of ex vivo cultures of latently infected WT TG. (Top) CD8$^+$ T cell in contact with two neurons polarizes TCR and GrB toward lower neuron only. (Bottom) CD8$^+$ T cells not contacting targets show no TCR or GrB polarization. DIC, differential interference contrast. (**C** to **E**) Imaging of ex vivo latently infected WT TG cultured 24 to 48 hours with WT gB-CD8, which prevent reactivation from latency. (**C**) Two representative interactions between NeuN$^+$ neurons and CD8$^+$ T cells with CD107a polarized toward the junction with a neuron lacking activated caspases (VAD-FMK$^-$). (Interaction 2, bottom) The plane demarcated by the line between the cells (left) is shown en face (right), demonstrating an apparent secretory domain of an immunological synapse. (**D**) 10% ethanol–induced neuronal caspase activation. (**E**) CD8$^+$ T cells contacting NeuN$^-$ noninfected fibroblastlike cells (left) or not contacting cells (right) lack surface CD107a expression. Data are representative of 11 TG from three separate experiments.

Fig. 2. Pfn and GrB are required to maintain HSV-1 neuronal latency in vivo and in ex vivo ganglia cultures. (**A**) DNA was extracted from individual TG at designated times, and HSV-1 genome copies were determined by quantitative real-time polymerase chain reaction (horizontal bar denotes mean). Data pooled from at least two independent experiments per time point. *$P \leq 0.05$, as calculated by analysis of variance with Bonferroni posttest. (**B**) 34 to 41 dpi TG were dispersed and cultured at one-fifth TG equivalents per well. HSV-1 reactivation was indicated by the presence of infectious virus in serially sampled supernatants, as assessed by plaque assay. Pooled data from three independent experiments presented as mean ± SEM (represented by error bars). *$P = 0.0009$ and **$P = 0.0002$, as calculated by survival curve analysis (log-rank test). (**C**) 14 dpi TG were dispersed and cultured at 1 TG per well in medium alone or in medium supplemented with 1000 U/ml rIFN-γ, 100 μg/ml anti-CD8α monoclonal antibody (α-CD8α mAb), or 20 μg/ml anti–IFN-γ monoclonal antibody (α-IFN-γ mAb). Reactivation was assessed, analyzed, and presented as in (B). $n = 10$ TG per condition. Data are representative of two independent experiments. *$P = 0.0108$ and **$P = 0.0049$. Error bars indicate SEM.
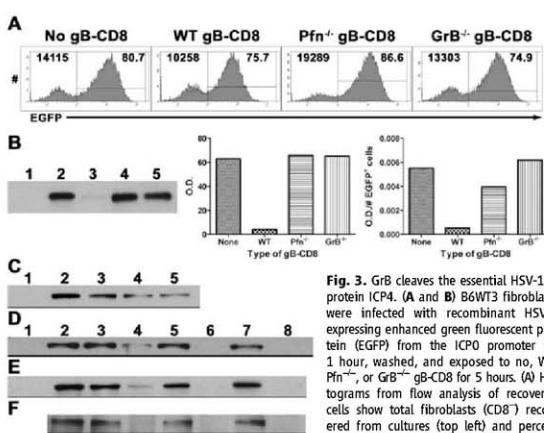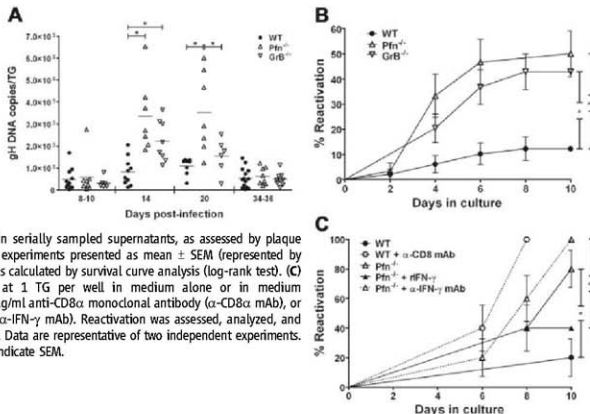
Fig. 3. GrB cleaves the essential HSV-1 IE protein ICP4. (**A** and **B**) B6WT3 fibroblasts were infected with recombinant HSV-1 expressing enhanced green fluorescent protein (EGFP) from the ICP0 promoter for 1 hour, washed, and exposed to no, WT, Pfn$^{-/-}$, or GrB$^{-/-}$ gB-CD8 for 5 hours. (A) Histograms from flow analysis of recovered cells show total fibroblasts (CD8$^-$) recovered from cultures (top left) and percent infected (EGFP$^+$; top right). (B) Lysates of recovered cells were subjected to Western blot for ICP4 (lane 1, noninfected; lanes 2 to 5, infected with No, WT, Pfn$^{-/-}$, or GrB$^{-/-}$ gB-CD8, respectively). Bar graphs show optical density (O.D.) readings that were not adjusted (middle) or adjusted for the number of infected fibroblasts recovered from each culture (right). (**C**) Lysates from 293T cells that were either nontransfected (lane 1) or transfected with an ICP4-expressing plasmid (lanes 2 to 5) and exposed to different concentrations of recombinant GrB (lanes 1 to 5: 0, 0, 25, 50, 100 nM GrB, respectively) for 1 hour at 37°C. Lysates from 293T cells transfected with an ICP4-expressing plasmid (**D**), lysates from B6WT3 cells infected with HSV-1 (**E**), or ICP4 immunoprecipitated from HSV-1-infected fibroblasts (**F**) were exposed to 0 (lanes 1 to 3, 5, and 7) or 100 nM GrB (lanes 4, 6, and 8) for varying times at 37°C (lanes 1 to 2, 0 hours; lanes 3 to 4, 1 hour; lanes 5 to 6, 2 hours; lanes 7 to 8, 3 hours). Lane 1 contains nontransfected 293T cell lysate (D) or noninfected fibroblast lysate (E).

HSV-1–infected fibroblasts to recombinant human GrB (Fig. 3, C to F). Demonstration of GrB and ICP4 colocalization within infected cells and the exact cellular compartment where cleavage occurs requires further investigation.

Understanding the mechanisms that control HSV-1 latency is of paramount importance because HSV-induced pathology is associated with viral shedding after reactivation from latency in sensory ganglia. Viral microRNAs (miRNAs) expressed during latency can inhibit production of multiple HSV-1 immediate early proteins, including ICP4 (24). Such mechanisms, in addition to epigenetic modifications of the viral genome (25), probably contribute to the stable latency that appears to exist in the vast majority of latently infected neurons. We propose that at any given time, some neurons escape these control mechanisms and require protection from HSV-specific CD8$^+$ T cells (26). It is these neurons that might represent those most likely to reactivate in vivo.

Our findings resolve the paradox that potentially cytotoxic CD8$^+$ T cells surround apparently healthy HSV-1 latently infected neurons in ganglia of both mice and humans. CD8$^+$ T cell lytic granules can block the HSV-1 life cycle through a nonlytic mechanism, and GrB can directly cleave ICP4, a viral protein required for efficient transcription of early and late viral genes. The fact that gB (a gene product expressed at low levels before viral DNA synthesis) is detectable by gB-CD8 within 2 hours of infection (27) would permit ICP4 degradation very early in the viral life cycle during both lytic infection and reactivation from latency. Blocking reactivation at this early point would prevent replication of viral DNA, consistent with our observation of a dramatic increase in viral genome

copy number in TG of GrB$^{-/-}$ or Pfn$^{-/-}$ mice. This mechanism might be particularly efficient during attempted HSV-1 reactivation events where ICP4 expression has escaped repression by viral miRNAs and host neuron epigenetic modifications. Thus, we propose a tripartite relation in which HSV-1 latency is maintained through the activity of the virus, host neuron, and contiguous CD8$^+$ T cells permitting viral persistence with neuronal survival (fig. S7).

References and Notes

1. D. Theil *et al.*, *Am. J. Pathol.* **163**, 2179 (2003).
2. K. Hufner *et al.*, *J. Neuropathol. Exp. Neurol.* **65**, 1022 (2006).
3. G. M. Verjans *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 3496 (2007).
4. T. Derfuss *et al.*, *Brain Pathol.* **17**, 389 (2007).
5. A. Simmons, D. C. Tscharke, *J. Exp. Med.* **175**, 1337 (1992).
6. E. M. Cantin, D. R. Hinton, J. Chen, H. Openshaw, *J. Virol.* **69**, 4898 (1995).
7. C. Shimeld *et al.*, *J. Neuroimmunol.* **61**, 7 (1995).
8. T. Liu, Q. Tang, R. L. Hendricks, *J. Virol.* **70**, 264 (1996).
9. K. M. Khanna, R. H. Bonneau, P. R. Kinchington, R. L. Hendricks, *Immunity* **18**, 593 (2003).
10. T. Liu *et al.*, *J. Exp. Med.* **191**, 1459 (2000).
11. M. L. Freeman, B. S. Sheridan, R. H. Bonneau, R. L. Hendricks, *J. Immunol.* **179**, 322 (2007).
12. K. D. Croen *et al.*, *N. Engl. J. Med.* **317**, 1427 (1987).
13. W. G. Stroop, D. C. Schaefer, *Acta Neuropathol.* **74**, 124 (1987).
14. T. Liu, K. M. Khanna, B. N. Carriere, R. L. Hendricks, *J. Virol.* **75**, 11178 (2001).
15. V. Decman, P. R. Kinchington, S. A. Harvey, R. L. Hendricks, *J. Virol.* **79**, 10339 (2005).
16. Materials and methods are available as supporting material on *Science* Online.
17. W. G. Telford, A. Komoriya, B. Z. Packard, *Cytometry* **47**, 81 (2002).
18. G.-C. Perng *et al.*, *Science* **287**, 1500 (2000).
19. Y. Hoshino, L. Pesnicak, J. I. Cohen, S. E. Straus, *J. Virol.* **81**, 8157 (2007).
20. R. A. Pereira, M. M. Simon, A. Simmons, *J. Virol.* **74**, 1029 (2000).
21. F. Andrade *et al.*, *EMBO J.* **26**, 2148 (2007).
22. C. Backes *et al.*, *Nucleic Acids Res.* **33**, W208 (2005).
23. N. A. DeLuca, A. M. McCarthy, P. A. Schaffer, *J. Virol.* **56**, 558 (1985).
24. J. L. Umbach *et al.*, *Nature* **454**, 780 (2008).
25. D. M. Knipe, A. Cliffe, *Nat. Rev. Microbiol.* **6**, 211 (2008).
26. B. S. Sheridan, J. E. Knickelbein, R. L. Hendricks, *Expert Opin. Biol. Ther.* **7**, 1323 (2007).
27. S. N. Mueller *et al.*, *J. Virol.* **77**, 2445 (2003).
28. We thank K. Lathrop and J. Karlsson for assistance with microscopy and preparation of figures and N. Zurowski for assistance with flow cytometry. We have no conflicting financial interests. This work was supported by NIH grants F30NS061471 (J.E.K.), R01EY05945 (R.L.H.), R01EY015291 (P.R.K.), and P30EY08098 (R.L.H.); a Research to Prevent Blindness Medical Student Eye Research Fellowship (J.E.K.); and unrestricted grants from Research to Prevent Blindness and the Eye and Ear Foundation of Pittsburgh (R.L.H.).

CTLA-4 Control over Foxp3$^+$ Regulatory T Cell Function

Kajsa Wing,[1]* Yasushi Onishi,[1,2] Paz Prieto-Martin,[1] Tomoyuki Yamaguchi,[1] Makoto Miyara,[1] Zoltan Fehervari,[1] Takashi Nomura,[1] Shimon Sakaguchi[1,3,4]†

Naturally occurring Foxp3$^+$CD4$^+$ regulatory T cells (Tregs) are essential for maintaining immunological self-tolerance and immune homeostasis. Here, we show that a specific deficiency of cytotoxic T lymphocyte antigen 4 (CTLA-4) in Tregs results in spontaneous development of systemic lymphoproliferation, fatal T cell–mediated autoimmune disease, and hyperproduction of immunoglobulin in mice, and it also produces potent tumor immunity. Treg-specific CTLA-4 deficiency impairs in vivo and in vitro suppressive function of Tregs—in particular, Treg-mediated down-regulation of CD80 and CD86 expression on dendritic cells. Thus, natural Tregs may critically require CTLA-4 to suppress immune responses by affecting the potency of antigen-presenting cells to activate other T cells.

N aturally occurring CD25$^+$CD4$^+$ regulatory T cells (Tregs), which specifically express the transcription factor Foxp3, suppress aberrant immune responses, including autoimmune diseases and allergy (*1*). Furthermore, reduction or expansion of Tregs can be exploited to provoke effective tumor immunity or transplantation tolerance, respectively. Two cardinal features of Foxp3$^+$ Tregs are that they constitutively express cytotoxic T lymphocyte antigen 4 (CTLA-4), which only happens after activation in other T cell subsets (*2–4*), and that Foxp3 controls the expression of CTLA-4 in Tregs (*5–9*). CTLA-4 is a potent nega-

tive regulator of T cell immune responses, as illustrated by CTLA-4 knockout (KO) mice, which die prematurely from multiorgan inflammation (*10*, *11*). The polymorphism of the CTLA-4 gene contributes substantially to the genetic susceptibility to autoimmune diseases such as type 1 diabetes (*12*). Moreover, autoimmunity, inflammatory bowel disease, and tumor immunity can be elicited by blocking CTLA-4 with a specific antibody (*3*, *4*, *13–15*). Yet the manner in which CTLA-4 negatively controls immune responses is controversial (*16*). CTLA-4 expressed by activated effector T cells may mediate a negative signal that attenuates their activation. Alternatively, but not exclusively, Foxp3$^+$ Tregs may require CTLA-4 for their suppressive function. By specifically deleting the CTLA-4 gene in Foxp3$^+$ Tregs in mice, we have attempted to determine the role of CTLA-4 for the maintenance of self-tolerance and immune homeostasis.

We generated BALB/c mice expressing Cre under the control of the Foxp3 promoter—hereafter called FIC (Fox-IRES-Cre) mice—and BALB/c mice expressing a floxed CTLA-4 gene (CTLA-4$^{fl/fl}$) [supporting online material (SOM) text and fig. S1] (*17*). Compared with BALB/c wild-type

(WT) mice, FIC mice expressed Foxp3 protein at slightly lower levels whereas CTLA-4$^{fl/fl}$ mice expressed equivalent levels of CTLA-4 (Fig. 1A). To assess the specificity of Cre expression, FIC mice were crossed with CAG reporter mice (CAG mice), which express enhanced green fluorescent protein (EGFP) only in Cre$^+$ cells (*18*). EGFP expression was confined to ~15% of CD4$^+$ T cells and ~1.5% of CD8$^+$ T cells (Fig. 1B). The vast majority of EGFP$^+$CD4$^+$ T cells in adult FIC$^{+/+}$ CAG mice were Foxp3$^+$ (97.1 ± 1.2%, $n = 4$ mice), indicating that Foxp3 expression is stable once the gene is turned on and Cre expression is not leaky in Foxp3$^−$ cells (Fig. 1C). On the basis of this specific expression of Cre in Foxp3$^+$ Tregs, we generated CTLA-4 conditional KO (CKO) mice by crossing FIC and CTLA-4$^{fl/fl}$ mice. CTLA-4 was specifically deleted in CD4$^+$Foxp3$^+$ T cells, as compared with FIC$^{+/+}$ WT or full CTLA-4 KO mice (Fig. 1D). CKO mice even harbored a higher frequency of CTLA-4–expressing CD4$^+$Foxp3$^+$ cells than did WT littermates (Fig. 1E). Whereas KO mice became moribund at ~20 days of age (*10*, *11*), CKO mice remained apparently unaffected until ~7 weeks of age, when they rapidly became inactive and began to develop general edema that was frequently accompanied by ascites (Fig. 1F). Thus, CTLA-4 deficiency in Tregs alone suffices to cause fatal disease, whereas the additional CTLA-4 deficiency in non-Treg cells enhances the disease. Yet, CTLA-4 expression in activated effector T cells per se is insufficient to prevent it.

Pathological analysis of CKO mice revealed splenomegaly and lymphadenopathy, which was reflected in increased cell numbers (Fig. 2, A and B). The proportion of CD4$^+$ T cells was unaltered, whereas CD8$^+$ T cells were decreased (Fig. 2C). Cardiomegaly and congestion of the liver was macroscopically evident in the terminal stage of every case. In affected hearts, mononuclear cells densely infiltrated into the myocardium and destroyed myocytes (Fig. 2, D to G), indicating that the plausible cause of sudden death in CKO mice is

[1]Department of Experimental Pathology, Institute for Frontier Medical Sciences, Kyoto University, Kyoto 606-8507, Japan. [2]Department of Rheumatology and Haematology, Tohoku University Graduate School of Medicine, Sendai 980-8574, Japan. [3]Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, Kawaguchi 332-0012, Japan. [4]Laboratory of Experimental Immunology, World Premier International Immunology Frontier Research Center, Osaka University, Suita 565-0871, Japan.

*Present address: Department of Medical Inflammation Research, Karolinska Institute, Stockholm 17177, Sweden.
†To whom correspondence should be addressed. E-mail: shimon@frontier.kyoto-u.ac.jp

heart failure due to severe myocarditis (19). In addition, CKO mice possessed focal lymphocyte infiltrations in lung and salivary gland and suffered from gastritis with various degrees of destruction of gastric parietal cells and chief cells. Antiparietal autoantibodies were readily detected in the sera of CKO mice and a proportion of FIC[+/+] mice, in which the lower expression of Foxp3 in Tregs (Fig. 1A) might somehow affect Treg function (20) (Fig. 2, H to N, and SOM text). Myocarditis and gastritis in CKO mice (and gastritis in FIC mice) could be adoptively transferred with splenocytes and purified CD4[+] T cells into T cell–deficient BALB/c athymic nude (nu/nu) mice, indicating that these autoimmune conditions were both T cell–mediated (Fig. 2O and fig. S2). Furthermore, CKO mice developed several hundred–fold and threefold higher levels of serum immunoglobulin E (IgE) and immunoglobulin G (IgG), respectively, than the levels in FIC or WT mice (Fig. 2, P and Q).

Costaining of intracellular cytokines and Foxp3 revealed an increased frequency of interleukin-2 (IL-2)–, IL-4–, and IFN-γ–producing Foxp3 CD4[+] cells in both the spleen and lymph node (LN) of diseased CKO and KO mice (Fig. 2R and fig. S3). IL-17–secreting (Th17) cells increased in KO but not CKO mice, suggesting that Th17 cells might contribute to the rapid disease progression in the former. Thus, CTLA-4–deficient Tregs fail to control the spontaneous activation of other T cells and their differentiation into Th1 and Th2 lineage cells that mediate autoimmune disease and allergy.

We next tested whether Treg-specific CTLA-4 deficiency also influenced the potency of tumor immunity. BALB/c nu/nu mice were reconstituted with splenocytes from CKO or control FIC mice containing equivalent numbers of T cells and inoculated with BALB/c-derived RL♂1 leukemia cells (21). All recipients of FIC splenocytes died of tumor progression within a month. In contrast, recipients of CKO splenocytes halted the tumor growth, with the majority surviving the 6-week observation period, during which 60% of them completely rejected the tumor (Fig. 3A). As previously shown (21), transfer of BALB/c splenocytes after depletion of CD25[+] T cells led to the rejection of RL♂1 leukemia cells in nu/nu mice. In this setting, FIC Tregs cotransferred with CD25 T cells suppressed tumor rejection, whereas CKO Tregs did not (Fig. 3B). Thus, Treg-specific CTLA-4 deficiency affects in vivo Treg suppressive function, leading to enhanced tumor immunity.

We next explored the possibility that CTLA-4 deficiency might impair the generation, survival, or suppressive function of Foxp3[+] Tregs. CKO mice exhibited no significant alteration in number or composition of CD4[+] and CD8[+] thymocytes (Fig. 4A). The majority of Foxp3[+] WT thymocytes expressed CTLA-4, whereas Foxp3[+] CKO thymocytes contained a mix of CTLA-4[+] and CTLA-4 cells in both the CD4–single positive and CD4/CD8–double positive compartments (Fig. 4A). Because the CTLA-4 gene is deleted only after Foxp3 is expressed, CTLA-4 is either up-regulated before Foxp3 expression in CKO mice or it may take some time for the Cre protein to accumulate in Foxp3[+] cells, meanwhile allowing the expression of CTLA-4. The frequency of Foxp3[+] thymocytes was not significantly changed between CKO and WT mice, whereas the number of Foxp3[+] and Foxp3 T cells in the spleen and LNs increased enormously by active proliferation (Fig. 4B, figs. S4 and S5, and SOM text). Thus, Foxp3-inducible CTLA-4 deficiency minimally alters thymic selection of Tregs and probably triggers immunological diseases through affecting Treg function in the periphery.

Because Foxp3 is encoded by the X chromosome, female nonautoimmune FIC[+/+] CTLA-4[fl/fl] mice are a mosaic for CTLA-4–intact and –deficient Tregs. They harbored equal numbers of CTLA-4[+] and CTLA-4 Foxp3[+] T cells, indicating that both populations equally survive in physiological non-inflammatory conditions (Fig. 4C). Furthermore, when CTLA-4–deficient or –intact Tregs were transferred to nu/nu both, both populations showed a similar degree of homeostatic proliferation, and neither one caused autoimmunity (fig. S6). CTLA-4–deficient Foxp3[+] Tregs were as poor at producing pro-inflammatory cytokines as were their WT or FIC counterparts (fig. S3). Taken together, CTLA-4 deficiency, per se, does not affect the survival of Tregs or render them pathogenic.

Phenotypically, CTLA-4–deficient naïve Tregs in FIC[+/+] CTLA-4[fl/fl] females normally expressed typical Treg markers including CD44, CD103, glucocorticoid-induced tumor necrosis factor receptor, latency-associated peptide, and intracellular IL-10 (Fig. 4D and fig. S7). The comparatively higher expression of these molecules by Tregs from CKO mice is presumably secondary to ongoing inflammation in CKO mice, as illustrated by an activated phenotype of their Foxp3 non-Treg cells.

CTLA-4–deficient Tregs, whether naïve from FIC[+/+] CTLA-4[fl/fl] CAG females or activated from CKO mice, had diminished suppressive capacity compared with CTLA-4–intact Tregs in cultures of carboxyfluorescein diacetate succinimidyl ester (CFSE)–labeled responder T cells (Tresp) in the presence of splenic CD11c[+] dendritic cells (DCs) and anti-CD3 monoclonal antibody (mAb), as assessed by the percentage and number of CFSE-diluting (i.e., divided) Tresp (Fig. 4E, figs. S8 and S9, and SOM text). Moreover, CKO Tregs clearly failed to suppress allo-reactive Tresp proliferation, even at high Treg/Tresp ratios (Fig. 4F). FIC or WT Tregs, whether cultured alone or together with Tresp cells, specifically hampered up-regulation of the expression of CD80 and CD86, but not CD40 and major histocompatability complex class II, in DCs (22–26). In contrast, CKO Tregs failed to exert this effect (Fig. 4G, figs. S10 to S13, and SOM text). Activated FIC Tregs (but

Fig. 1. Specific deletion of CTLA-4 expression in Foxp3[+] T cells results in fatal disease. (A) Flow cytometric analysis of intracellular Foxp3 (left) and CTLA-4 (right) in freshly isolated LN CD4[+] T cells from male FIC, CTLA-4[fl/fl], or BALB/c WT mice. (B) EGFP expression in CD4[+] or CD8[+] T cells derived from male FIC-CAG mice. (C) Sorted CD4[+]EGFP[+] cells in FIC-CAG mice were stained for Foxp3. (D) CTLA-4 and Foxp3 expression in LN CD4[+] T cells from BALB/c WT, CKO, or KO mice. (E) Frequency of CTLA-4–expressing CD4[+]Foxp3[−] T cells in CKO and normal littermates (n = 5). (F) Survival of KO and CKO mice as compared with normal littermates. Data represent three or more independent experiments. Vertical bars indicate SEM.
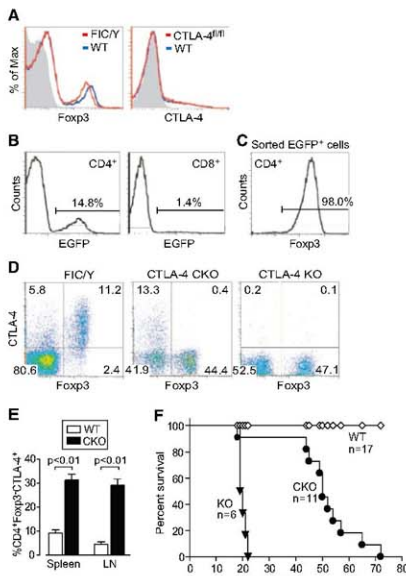
Fig. 2. Autoimmune disease and hyperproduction of IgE in CKO mice. (**A**) Splenomegaly and lymphadenopathy in a CKO and a WT littermate. Lymphocyte numbers (**B**) and frequencies of T cell subsets (**C**) in spleens of 6- to 10-week-old CKO and WT littermates ($n = 11$ to 13). (**D**) The heart of a CKO (left) and a FIC$^{+/+}$CTLA-4$^{fl/fl}$ mouse (right). Histology (hematoxylin and eosin staining) of the heart of a CKO [(**E** and **F**) ×50 and ×200, respectively] and a FIC mouse [(**G**) ×50; inset, ×200]. Histology of the stomach [(**H** and **I**) ×100], lung [(**J** and **K**) ×100], and salivary gland [(**L** and **M**) ×50] of a CKO [(**H**), (**J**), and (**L**)] and a FIC mouse [(**I**), (**K**), and (**M**)]. Serological and histological development of gastritis in WT, FIC$^{+/+}$, and CKO mice (**N**), and BALB/c nu/nu mice 7 weeks after cell transfer from CKO or FIC$^{+/+}$ mice (**O**). Gastric lesions were histologically graded as 2 (black circle), 1 (gray circle), and 0 (open circle) (19). Serum concentrations of IgE (**P**) and IgG (**Q**) in indicated groups of mice. (**R**) Frequencies of cytokine-producing cells among CD4$^+$Foxp3$^-$ splenocytes of 6- to 9-week-old CKO, 16- to 20-day-old KO, or normal littermates ($n = 5$ to 6). Error bars indicate SEM.
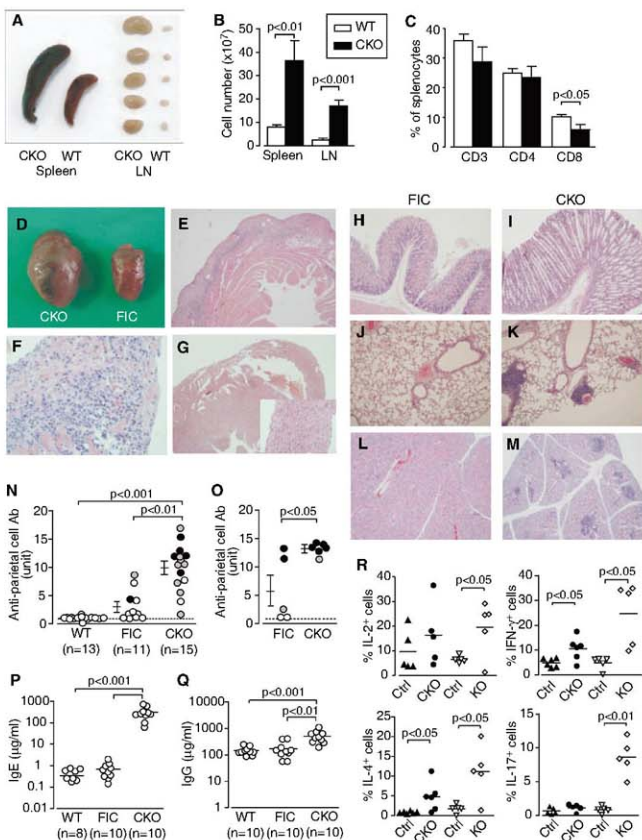
not activated CKO Tregs also reduced the expression of CD86-GFP fusion protein retrovirally expressed in L-cells, a fibroblast cell line (Fig. 4H). This indicates that Treg-dependant modulation of CD86 expression on DCs is at least partly due to down-regulation of the expression and not masking of the molecule by soluble CTLA-4. Taken together, Treg-mediated CD80/CD86 downregulation may limit the activation of naive T cells via CD28, resulting in specific immune suppression and tolerance.

Thus, CTLA-4 expressed in Foxp3$^+$ Tregs is critically required for their in vivo and in vitro suppression, which is mediated at least in part by

Fig. 3. Treg-specific CTLA-4 deficiency promotes tumor immunity. (**A**) BALB/c nu/nu mice received 3×10^7 splenocytes from FIC or CKO mice, followed by intradermal inoculation of 1.5×10^5 RL♂1 leukemia cells. Crosses indicate death due to tumor growth. (**B**) BALB/c CD25$^-$ cells (1.5×10^7) were cotransferred with

3.8×10^5 CD25^{high}CD4$^+$ T cells from CKO or FIC mice and inoculated with 1.5×10^5 RL♂1 cells ($n = 3$). Tumor diameters were measured every other day for 6 weeks. Mice were euthanized when tumor diameters exceeded 20 mm. Error bars indicate SEM.
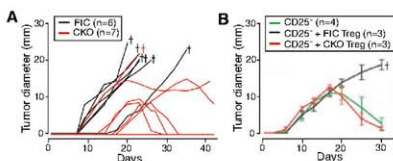
Fig. 4. CTLA-4–deficient Tregs develop and survive normally but have defective function. (**A**) Thymocyte expression of Foxp3 and CTLA-4 in a 2.5-week-old CKO or a WT littermate. SP, single positive; DP, double positive. (**B**) Frequency and number of CD4⁺Foxp3⁺ T cells in spleens and LNs of 6- to 8-week-old CKO or WT littermates (*n* = 7). (**C**) Foxp3 and CTLA-4 expression in splenic CD4⁺ T cells from a FIC⁺/⁻CTLA-4^fl/fl female mouse (left). Percentages of Foxp3⁺CTLA-4⁺ and Foxp3⁺CTLA-4⁻ T cells in each mouse (5 to 8 weeks of age) are connected (*n* = 8) (right). ns, not significant. (**D**) Expression of cell surface molecules on CD4⁺ LN cells from CKO, FIC, or FIC⁺/⁻CTLA-4^fl/fl mice. (**E**) CD25^high EGFP⁺ cells (naive CTLA-4⁻ Treg) and CD25^high EGFP⁻ cells (naive CTLA-4⁺ Treg) from FIC⁺/⁻CTLA-4^fl/fl CAG female mice and CD25^high CD4⁺ T cells from FIC mice (FIC Treg) were cocultured with CD25⁻CD4⁺ T cells (Tresp), anti-CD3 mAb, and live splenic DCs for 3 days. Percentages and numbers (in parentheses) of CFSE-diluting Tresp cultured at a 1:2 Treg-to-Tresp ratio (left). Numbers of CFSE-diluting Tresp cultured at graded ratios of Treg:Tresp (right). (**F**) Percentage and numbers of CFSE-labeled BALB/c Tresp cocultured with CKO or FIC Tregs and X-irradiated C57BL/6 splenocytes for 4 days at 1:1 Treg:Tresp ratio (left) and numbers at graded ratios (right). (**G**) CD80 and CD86 expression of live splenic DCs after a 2-day culture with Tresp, CD4⁺EGFP⁺ Tregs, or a mix thereof, and anti-CD3 mAb. Histograms show mean fluorescence intensity (MFI). (**H**) L-cells, expressing the Fc receptor, were retrovirally transduced to express CD86-EGFP fusion protein, cocultured with indicated Tregs and anti-CD3 mAb for 2 days, and assessed for GFP

level. Data in (A) and (D) to (H) represent three or more independent experiments. Error bars indicate SEM.
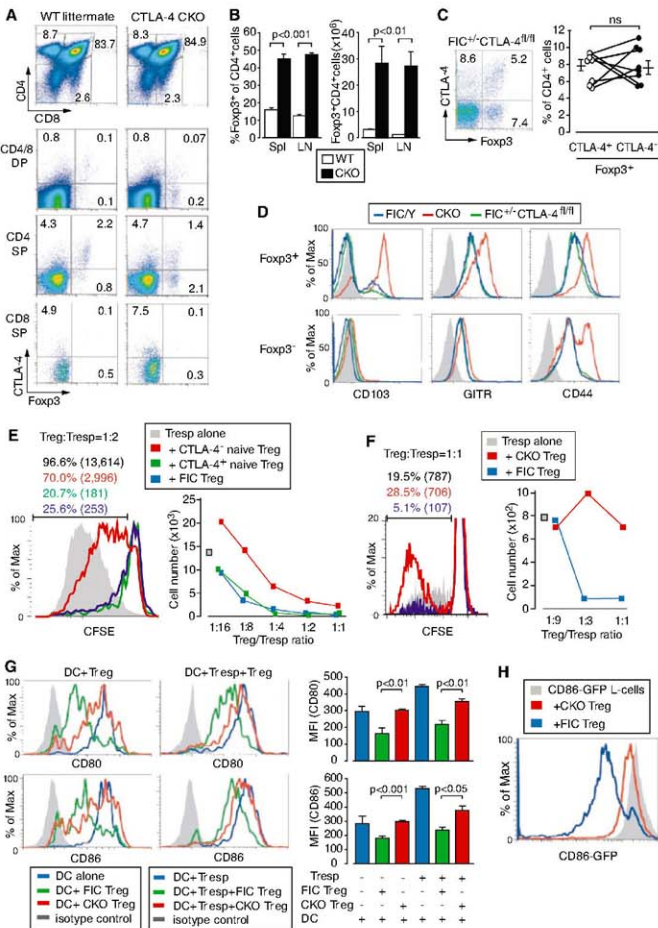
CTLA-4–dependent down-regulation of CD80 and CD86 on antigen presenting cells. Tregs probably use multiple suppressive mechanisms, and the importance of each one may vary depending on the environment and the context of immune responses (*1*). However, if the CTLA-4–mediated mechanism of suppression is defective, Tregs cannot sustain self-tolerance and immune homeostasis, even if other suppressive mechanisms become more active to compensate for the deficiency. Thus, CTLA-4 is a key molecular target for controlling Treg-suppressive function in both physiological and pathological immune responses including autoimmunity, allergy, and tumor immunity.

References and Notes
1. S. Sakaguchi *et al.*, *Cell* **133**, 775 (2008).
2. B. Salomon *et al.*, *Immunity* **12**, 431 (2000).
3. T. Takahashi *et al.*, *J. Exp. Med.* **192**, 303 (2000).
4. S. Read, V. Malmstrom, F. Powrie, *J. Exp. Med.* **192**, 295 (2000).
5. J. Hori, T. Nomura, S. Sakaguchi, *Science* **299**, 1057 (2003); published online 9 January 2003; 10.1126/science.1079490.

REPORTS

6. Y. Wu et al., Cell 126, 375 (2006).
7. A. Marson et al., Nature 445, 931 (2007).
8. Y. Zheng et al., Nature 445, 936 (2007).
9. M. Ono et al., Nature 446, 685 (2007).
10. P. Waterhouse et al., Science 270, 985 (1995).
11. E. A. Tivol et al., Immunity 3, 541 (1995).
12. The Wellcome Trust Case Control Consortium, Nature 447, 661 (2007).
13. D. R. Leach, M. F. Krummel, J. P. Allison, Science 271, 1734 (1996).
14. G. Q. Phan et al., Proc. Natl. Acad. Sci. U.S.A. 100, 8372 (2003).
15. S. Read et al., J. Immunol. 177, 4376 (2006).
16. D. M. Sansom, L. S. Walker, Immunol. Rev. 212, 131 (2006).
17. Materials and methods are available as supporting material on Science Online.
18. S. Kawamoto et al., FEBS Lett. 470, 263 (2000).
19. M. Ono, J. Shimizu, Y. Miyachi, S. Sakaguchi, J. Immunol. 176, 4748 (2006).
20. Y. Y. Wan, R. A. Flavell, Nature 445, 766 (2007).
21. J. Shimizu, S. Yamazaki, S. Sakaguchi, J. Immunol. 163, 5211 (1999).
22. L. Cederbom, H. Hall, F. Ivars, Eur. J. Immunol. 30, 1538 (2000).
23. C. Oderup, L. Cederbom, A. Makowska, C. M. Cilio, F. Ivars, Immunology 118, 240 (2006).
24. S. Yamazaki, K. Inaba, K. V. Tarbell, R. M. Steinman, Immunol. Rev. 212, 314 (2006).
25. R. J. DiPaolo et al., J. Immunol. 179, 4685 (2007).
26. Y. Onishi et al., Proc. Natl. Acad. Sci. U.S.A. 105, 10113 (2008).
27. We thank M. Ono for discussion and R. Ishii and M. Matsushita for technical assistance. This work was supported by Grants-in-Aid from the Ministry of Education, Sports and Culture of Japan, Japan Science and Technology Agency. Z.F. was a Japan Society for the Promotion of Science fellow, and K.W. was granted a fellowship by Astra-Zeneca, Loughborough, UK.

Environmental Genomics Reveals a Single-Species Ecosystem Deep Within Earth

Dylan Chivian,[1,2*] Eoin L. Brodie,[2,3] Eric J. Alm,[2,4] David E. Culley,[5] Paramvir S. Dehal,[1,2] Todd Z. DeSantis,[2,3] Thomas M. Gihring,[6] Alla Lapidus,[7] Li-Hung Lin,[8] Stephen R. Lowry,[9] Duane P. Moser,[9] Paul M. Richardson,[7] Gordon Southam,[10] Greg Wanger,[10] Lisa M. Pratt,[11,12] Gary L. Andersen,[2,3] Terry C. Hazen,[2,3,12] Fred J. Brockman,[13] Adam P. Arkin,[1,2,14] Tullis C. Onstott[12,15]

DNA from low-biodiversity fracture water collected at 2.8-kilometer depth in a South African gold mine was sequenced and assembled into a single, complete genome. This bacterium, Candidatus Desulforudis audaxviator, composes >99.9% of the microorganisms inhabiting the fluid phase of this particular fracture. Its genome indicates a motile, sporulating, sulfate-reducing, chemoautotrophic thermophile that can fix its own nitrogen and carbon by using machinery shared with archaea. Candidatus Desulforudis audaxviator is capable of an independent life-style well suited to long-term isolation from the photosphere deep within Earth's crust and offers an example of a natural ecosystem that appears to have its biological component entirely encoded within a single genome.

A more complete picture of life on, and even in, Earth has recently become possible by extracting and sequencing DNA from an environmental sample, a process called environmental genomics or metagenomics (1–8). This approach allows us to identify members of microbial communities and to characterize the abilities of the dominant members even when isolation of those organisms has proven intractable. However, with a few exceptions (5, 7), assembling complete or even near-complete genomes for a substantial portion of the member species is usually hampered by the complexity of natural microbial communities.

In addition to elevated temperatures and a lack of O_2, conditions within Earth's crust at depths >1 km are fundamentally different from those of the surface and deep ocean environments. Severe nutrient limitation is believed to result in cell doubling times ranging from 100s to 1000s of years (9–11), and as a result subsurface microorganisms might be expected to reduce their reproductive burden and exhibit the streamlined genomes of specialists or spend most of their time in a state of semi-senescence, waiting for the return of favorable conditions.

Such microorganisms are of particular interest because they permit insight into a mode of life independent of the photosphere.

One bacterium belonging to the Firmicutes phylum (Fig. 1A), which we herein name Candidatus Desulforudis audaxviator, is prominent in small subunit (SSU or 16S) ribosomal RNA (rRNA) gene clone libraries (11–14) from almost all fracture fluids sampled to date from depths greater than 1.5 km across the Witwatersrand basin (covering 150 km by 300 km near Johannesburg, South Africa). This bacterium was shown in a previous geochemical and 16S rRNA study (11) to dominate the indigenous microorganisms found in a fracture zone at 2.8 km below land surface at level 104 of the Mponeng mine (MP104). Although Lin et al. (11) discovered that this fracture zone contained the least-diverse natural free-living microbial community reported at that time, exceeding the ~80% dominance by the methanogenic archaeon IUA5/6 of a comparatively shallow subsurface community in Idaho (15) we were nonetheless surprised when the current environmental genomics study revealed only one species was actually present within the fracture fluid. Furthermore, we found that the

genome of this organism appeared to possess all of the metabolic capabilities necessary for an independent life-style. This gene complement was consistent with the previous geochemical and thermodynamic analyses at the ambient ~60°C temperature and pH of 9.3, which indicated radiolytically generated chemical species as providing the energy and nutrients to the system (11), with formate and H_2 as possessing the greatest potential among candidate electron donors, and sulfate (SO_4^{2-}) reduction as the dominant electron-accepting process (11).

DNA was extracted from ~5600 liters of filtered fracture water by using a protocol that has been demonstrated to be effective on a broad range of bacterial and archaeal species, including recalcitrant organisms (16). A single, complete, 2.35-megabase pair (Mbp) genome was assembled with a combination of shotgun Sanger sequencing and 454 pyrosequencing (16). Similar to other studies that obtained near-complete consensus genomes from environmental samples (5, 17), heterogeneity in the population of the dominant species as measured with single-nucleotide polymorphisms (SNP) was quite low, showing only 32 positions with a SNP observed

[1]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [2]Virtual Institute for Microbial Stress and Survival, Berkeley, CA 94720, USA. [3]Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [4]Departments of Biological and Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [5]Energy and Efficiency Technology Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA. [6]Department of Oceanography, Florida State University, Tallahassee, FL 32306, USA. [7]Genomic Technology Program, U.S. Department of Energy (DOE) Joint Genomics Institute, Berkeley, CA 94598, USA. [8]Department of Geosciences, National Taiwan University, Taipei 106, Taiwan. [9]Division of Earth and Ecosystem Sciences, Desert Research Institute, Las Vegas, NV 89119, USA. [10]Department of Earth Sciences, University of Western Ontario, London, ON N6A 5B7, Canada [11]Department of Geological Sciences, Indiana University, Bloomington, IN 47405, USA. [12]Indiana Princeton Tennesse Astrobiology Initiative (IPTAI), NASA Astrobiology Institute, Bloomington, IN 47405, USA. [13]Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA 99352, USA. [14]Department of Bioengineering, University of California, Berkeley, CA 94720, USA. [15]Department of Geosciences, Princeton University, Princeton, NJ 08544, USA.

*To whom correspondence should be addressed. E-mail: DCChivian@lbl.gov

A Phylogenetic classification

- Bacillus subtilis subsp. subtilis str. 168
- Thermoanaerobacter tengcongensis MB4
- Clostridium novyi NT
- Thermosinus carboxydivorans Nor1
- Symbiobacterium thermophilum IAM 14863
- Heliobacterium modesticaldum Ice1
- Desulfitobacterium hafniense Y51
- Moorella thermoacetica ATCC 39073
- Syntrophomonas wolfei subsp. wolfei str. Goettingen
- Carboxydothermus hydrogenoformans Z-2901
- Desulfotomaculum reducens MI-1
- Pelotomaculum thermopropionicum SI
- Candidatus Desulforudis audaxviator MP104C

● Bootstrap support > 90%
— 0.1 JTT distance

B SSU rRNA clone library (361 clones)

97.2%

- ■ Desulforudis audaxviator 99+ (96.1%)
- □ Likely Desulforudis audaxviator 98+ (1.1%)
- ■ α-Proteobacterium type #1, Sphingomonas-like #1 (1.4%)
- ■ β-Proteobacterium type #1, Aquabacterium #1 (1.1%)
- ■ β-Proteobacterium type #2, Burkholderia-group #1 (0.3%)

C Sanger metagenomic sequence (29180 microbial reads)

99.9%

- ■ Desulforudis audaxviator (29136 reads = 99.849%)
- □ Likely Desulforudis audaxviator (17 reads = 0.058%)
- ■ Other Bacteria and Archaea, including likely microbial contamination (27 reads = 0.093%)

D 454 metagenomic sequence (500130 microbial reads)

99.9%

- ■ Desulforudis audaxviator (499699 reads = 99.914%)
- □ Likely Desulforudis audaxviator (250 reads = 0.050%)
- ■ Other Bacteria and Archaea, including likely microbial contamination (181 reads = 0.036%)

Fig. 1. Phylogeny and population structure. (**A**) Phylogenetic placement of *D. audaxviator* based on protein sequences of universal protein families (table S3). High bootstrap value–supported nodes are indicated with circles. (**B**) Classifications of SSU rRNA gene clones from polymerase chain reaction amplification of filter extract (fig. S3). (**C**) Proportions of Sanger sequencing reads from shotgun clone library of filter extract. Reads classified as *D. audaxviator* by match to assembled genome or by match to sequenced organisms (table S6). (**D**) Proportions of 454 metagenomic reads directly from filter extract. Reads classified as *D. audaxviator* by match to assembled genome or by match to sequenced organisms (table S6).

more than once (table S7), suggesting a recent selective sweep or other population bottleneck.

The DNA recovered from the filter, assuming the capture of cells and extraction of DNA from those cells was indeed comprehensive, revealed that this genome represented the only species present in the fluid phase of the fracture. Of the ~0.1% of microbial reads not belonging to *D. audaxviator* (Fig. 1, C and D, and tables S5 and S6), about one-half represented clear laboratory contamination (table S6), the removal of which resulted in only 22 of 29,179 Sanger reads (0.075%) and 59 of 500,008 pyrosequencing reads (0.012%) that could be from other microorganisms. Despite precautions taken in collecting the sample, some of the trace reads could come from microbial contaminants in the mine. An upper-bound estimate of the contribution of any microorganism other than *D. audaxviator* to the community (table S6) offered at most only five Sanger reads (0.017%) corresponding to γ-Proteobacteria or at most nine pyrosequencing reads (0.0018%) corresponding to α-Proteobacteria, both of which are common in the mining water (*11*, *14*). Even if these Proteobacteria were not contaminants, it is unlikely that *D. audaxviator*, and indeed the functioning of the ecosystem, is metabolically dependent on organisms that would be outnumbered by about 5000 to 1 (or about 50,000 to 1 from the pyrosequencing data). However, we could not rule out the presence of organisms that might adhere to the surfaces of the fracture or that were smaller than the 0.2-μm filter pore that might play a role in the MP104 ecosystem, perhaps as reservoirs of genetic variation (*18*).

We analyzed the genome of *D. audaxviator* with use of MicrobesOnline (www.microbesonline. org) (*19*). If *D. audaxviator* is indeed the solitary resident of this habitat, then its genome should contain the complete genetic complement for

maintaining the biological component of the ecosystem, which would prohibit the extreme reduction of its genome. The genome (Table 1) at 2.35 Mbp was smaller than the 3 Mbp of its nearest sequenced relative, *Pelotomaculum thermopropionicum*. It contained 2157 predicted protein coding genes, more than found in streamlined free-living microorganisms, which typically have fewer than 2000 genes (*20*). We found all of the processes necessary for life encoded within the genome, including energy metabolism, carbon fixation, and nitrogen fixation.

Consistent with the thermodynamic evaluation (*11*) that SO_4^{2-} offers the most energetically favorable electron acceptor, the genome possesses the capacity for dissimilatory sulfate reduction (DSR) (Figs. 2 and 3 and table S13) with a gene repertoire like that of other SO_4^{2-}-reducing microorganisms (*21*). These genes are present in a set of operons (labeled SR1 to SR11 in Fig. 2) and include an extra copy of an archaeal-type sulfate adenylyltransferase (Sat) (fig. S5) and a H^+-translocating pyrophosphatase, both of which appear to be a consequence of horizontal gene transfer (HGT). High potential electrons probably enter primarily via the activity of a variety of hydrogenases acting upon H_2 (table S24).

Carbon assimilation may be from a variety of sources depending on local conditions. The genome contains sugar and amino acid transporters (Fig. 3 and table S20), suggesting that, at locations where biodensity is high, heterotrophic sources could be used, including recycling of dead cells. At MP104, where biodensity is low, carbon is fixed from inorganic sources. *D. audaxviator* appeared not to use the reverse TCA cycle (table S23) but did have all the machinery of the acetyl–coenzyme A (CoA) synthesis (Wood-Ljungdahl) pathway (*22*, *23*), which uses carbon monoxide dehydrogenase (CODH) for the assimilation of inorganic carbon (Figs. 2

Table 1. General features of the *D. audaxviator* genome.

Feature	Value
Genome size (bp)	2,349,476
G+C content (%)	60.9
Predicted protein coding genes (CDS/ORF)	2157
Genes without homology to other organisms (ORFans)	210
Pseudogenes derived from a protein coding gene	83
Average CDS/ORF length (bp)	910
Longest CDS/ORF length (bp)	5601
Percent of genome protein coding (%)	86.8
rRNA operons (16S-23s-5S)	2
Transfer RNAs (all amino acids represented, including selenocysteine)	45
Other nonprotein coding RNAs	7
CRISPR regions	2
Mobile elements (transposons/integrases)	
Gene groups	30
Gene count	83
Other phage-associated genes	18

and 3, fig. S7, and table S14). Entry of CO_2 substrate into the cell may be accomplished by its anionic species through a putative carbonate adenosine triphosphate (ATP)–binding cassette transporter or a putative bicarbonate/Na^+ symporter (Fig. 3 and table S20). Formate and CO may serve as alternate, more direct, carbon sources in other fractures when sufficiently abundant (table S2).

The ambient concentration of ammonia in the fracture water ([NH_3] + [NH_4^+] = ~100 μM) (*11*) appears sufficient for *D. audaxviator* (which has an ammonium transporter as well as glutamine
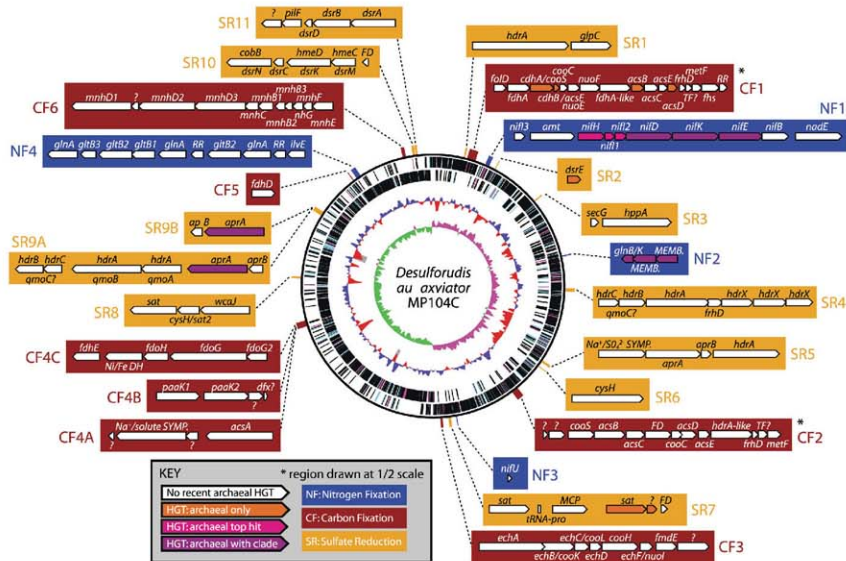
Fig. 2. Genome of *D. audaxviator*, with key genes highlighted. (Innermost ring) GC skew [average of (G-C)/(G+C) over 10,000 bases, plotted every 1000 bases]. Transition at the top (near dnaA) is origin of replication. (Second ring) G+C content [average of (G+C) over 10,000 bases, plotted every 1000 bases], with greater-than-average values (61%) in blue and below average in red. Below-average G+C regions that result from CRISPR sequences are indicated in gray. (Third and fourth rings) Predicted protein coding genes on each strand. Genes with homologs are found within

closest clade species [including open reading frame (ORF)an genes] are in cyan, genes that are found only within closest clade species and within archaea (resulting from horizontal transfer) in magenta, and all other genes in black. (Outer boxes) Genes of interest are shown around the ring as operons for sulfate reduction (SR), carbon fixation via acetyl-CoA synthesis pathway (CF), and nitrogen fixation (NF). Horizontally acquired genes shared with archaea specific to *D. audaxviator* and its nearest relatives are colored according to the key.

synthetase) to obtain its nitrogen from ammonia without resorting to an energetically costly nitrogenase conversion of N_2 to ammonia. Nonetheless, a nitrogenase is present in the genome (Fig. 2 and table S15) with a nifH subunit that is more similar to archaeal types, including high-temperature variants (24), than to the nitrogenase of *Desulfotomaculum reducens* (figs. S4 and S8). It may be that *D. audaxviator* is not always presented with sufficient amounts of ammonia, so the versatility provided by the horizontally acquired nitrogenase may have contributed substantially to the success of *D. audaxviator* in colonizing such habitats.

D. audaxviator shares other genes with archaea that may confer benefits in extreme environments. In addition to the unusual nitrogenase and sulfate adenylyltransferase, acquisitions by ancestors of *D. audaxviator* include a second CODH system (CODH1 in Fig. 2 and fig. S7), cobalamin biosynthesis protein CobN, and genes for the formation of gas vesicles. It also has two

clustered regularly interspaced short palindromic repeat (CRISPR) regions (table S12) that are used for viral defense (25) and that occur in the genome with adjacent CRISPR-associated genes (CAS), some of which are horizontally shared between *D. audaxviator* and archaea.

D. audaxviator's ability to colonize independently is also assisted by its possession of all of the amino acid synthesis pathways (table S21). Other factors that may confer fitness in this environment are the ability to form endospores (table S16) and the potential for it to grow in deeper, hotter conditions (table S9) than provided by MP104. *D. audaxviator* appears capable of sensing nutrients (table S19) in its environment and possesses flagellar genes (table S18) to permit motility along chemical gradients, such as those that occur at the mineral surfaces of the fracture (26). One ability that *D. audaxviator* is lacking is a complete system for oxygen resistance (table S25), suggesting the long-term isolation from O_2.

The MP104 fracture contains the simplest natural environmental microbial community yet described and has yielded a single, complete genome of an uncultured microorganism with the use of environmental genomics. *D. audaxviator*'s ability to reduce SO_4^{2-} grants access to the most energetically favorable electron acceptor in the fracture zones of the Witwatersrand basin (27). Additionally, inherited characteristics of *D. audaxviator*, such as motility, sporulation, and carbon fixation, have been complemented by horizontally acquired systems frequently found in archaea. These abilities have enabled *D. audaxviator* to colonize the deep subsurface, a process that, unlike surface habitats which permit more immediate access, has required fitness throughout the history of the colonization. This bold traveler (*audax viator*) has revealed a mode of life isolated from the photosphere, capturing all of the roles necessary for an independent life-style and showing that it is possible to encode the entire biological component of a simple ecosystem within a single genome.
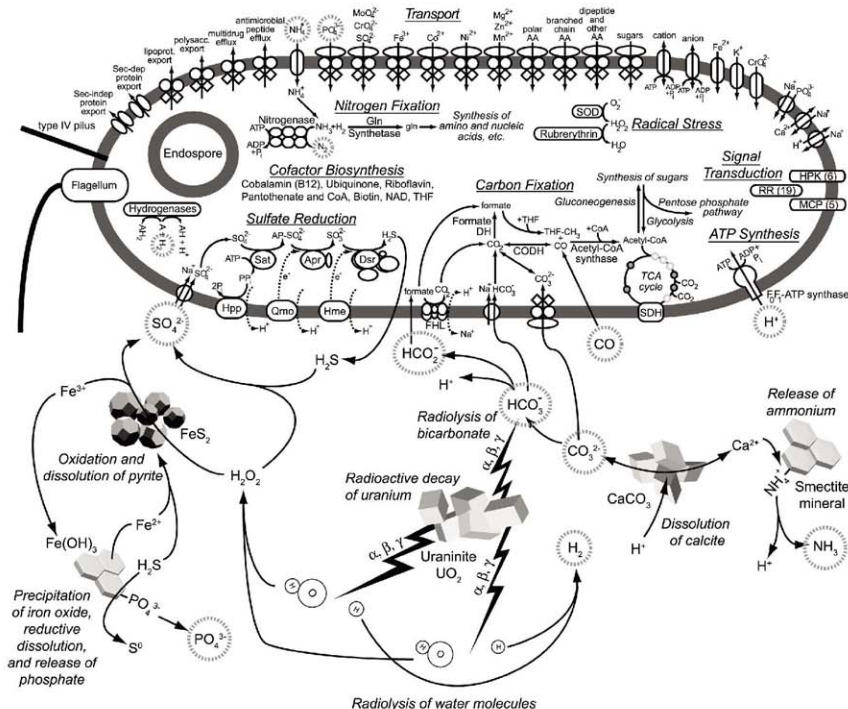
Fig. 3. Model of the single-species ecosystem at MP104. *D. audaxviator*'s machinery is shown in a cartoon representation, including pathways for sulfate reduction, nitrogen fixation, and carbon fixation. Signal transduction proteins are reported including the number found in parentheses (with MCP indicating methyl-accepting

chemotaxis proteins; HPK, histidine protein kinases; and RR, response regulators. Transporters include approximate substrates. Also shown are the radiolytically generated sources of energy and nutrients for the ecosystem, as detailed in Lin *et al.* (*11*), shown experimentally by Lefticariu *et al.* (*28*), and described in (*16*).

References and Notes

1. A. M. Deutschbauer, D. Chivian, A. P. Arkin, *Curr. Opin. Biotechnol.* **17**, 229 (2006).
2. O. Beja et al., *Environ. Microbiol.* **2**, 516 (2000).
3. M. R. Rondon et al., *Appl. Environ. Microbiol.* **66**, 2541 (2000).
4. J. C. Venter, *Science* **304**, 66 (2004); published online 4 March 2004 (10.1126/science.1093857).
5. G. W. Tyson et al., *Nature* **428**, 37 (2004).
6. S. G. Tringe, *Science* **308**, 554 (2005).
7. M. Strous et al., *Nature* **440**, 790 (2006).
8. D. B. Rusch et al., *PLoS Biol.* **5**, e77 (2007).
9. T. J. Phelps, E. M. Murphy, S. M. Pfiffer, D. C. White, *Microb. Ecol.* **28**, 335 (1994).
10. B. B. Jørgensen, S. D'Hondt, *Science* **314**, 932 (2006).
11. L. H. Lin et al., *Science* **314**, 479 (2006).
12. D. P. Moser et al., *Appl. Environ. Microbiol.* **71**, 8773 (2005).
13. D. P. Moser et al., *Geomicrobiol. J.* **20**, 517 (2003).
14. T. M. Gihring et al., *Geomicrobiol. J.* **23**, 415 (2006).
15. F. H. Chapelle et al., *Nature* **415**, 312 (2002).
16. Materials and methods are available as supporting material on *Science* Online.
17. V. Zverlov et al., *J. Bacteriol.* **187**, 2203 (2005).
18. M. L. Sogin et al., *Proc. Natl. Acad. Sci. U.S.A.* **103**, 12115 (2006).
19. E. J. Alm et al., *Genome Res.* **15**, 1015 (2005).
20. S. J. Giovannoni et al., *Science* **309**, 1242 (2005).
21. M. Mussmann et al., *J. Bacteriol.* **187**, 7126 (2005).
22. H. L. Drake, S. L. Daniel, *Res. Microbiol.* **155**, 869 (2005).
23. M. Wu et al., *PLoS Genet.* **1**, e65 (2005).
24. M. P. Mehta, J. A. Baross, *Science* **314**, 1783 (2006).
25. R. Barrangou et al., *Science* **315**, 1709 (2007).
26. G. Wanger, T. C. Onstott, G. Southam, *Geomicrobiol. J.* **23**, 443 (2006).
27. T. C. Onstott et al., *Geomicrobiol. J.* **23**, 369 (2006).
28. L. Lefticariu, L. M. Pratt, E. M. Ripley, *Geochim. Cosmochim. Acta* **70**, 4889 (2006).
29. We thank J. Banfield and G. Tyson for helpful discussion; J. Bruckner and B. Baker for assistance with microscopy; F. Warnecke for advice on 16S fluorescent in situ hybridization; T. Kieft, G. Zane, and the MicrobesOnline team (M. Price, K. Keller, and K. Huang) for advice; and D. Kershaw and colleagues at the Mponeng mine and AngloGold Ashanti Limited, RSA. This work was part of the

Supporting Online Material

www.sciencemag.org/cgi/content/full/322/5899/275/DC1
Materials and Methods
Figs. S1 to S8
Tables S1 to S26
References

Membrane Filtration

The Millex HPF line of syringe filters is now available with a new hydrophilic Teflon membrane that exhibits broad chemical compatibility and low extractability. The Millex HPF filter incorporates multiple media in its design: a graduated glass fiber prefilter and a membrane filter. This multilayer filter is resistant to clogging by viscous and high-particulate solutions, yet provides quantitative retention of fine particulates. The Teflon and nylon Millex HPF filters are available with 0.2-μm and 0.45-μm pore sizes, for filtering samples and mobile phases for ultrahigh performance liquid chromatography.

Millipore
For information 800-548-7853
www.millipore.com

Chemiluminescence Detection System

OptiBlaze ELISA femto-HRP is a sensitive chemiluminescence detection system for antibodies conjugated to horseradish peroxidase (HRP). Developed for luminometer-based applications, the chemiluminescent substrates are sensitive and stable. This system is able to detect low femtogram to picogram levels of enzyme, making it suitable for enzyme-linked immunosorbent assays (ELISAs). It is also available for detection of alkaline phosphatase-conjugated antibodies.

G-Biosciences/Genotech
For information 314-991-6034
www.GBiosciences.com

Chromatography Media

Two new high-performance affinity chromatography media enable purification under physiological conditions and mild elution, which means that the activity of the target protein is preserved. Designed for the purification of proteins tagged with maltose binding proteins, the small, evenly sized 34-μm Dextrin Sepharose High Performance beads ensure that proteins elute in narrow peaks, thus minimizing the need for further concentration steps. StrepTactin Sepharose High Performance is designed for Strep-tag II fusion proteins. Both media are available in prepacked columns.

GE Healthcare
For information 732-457-8149
www.gelifesciences.com/protein-purification

Membrane Protein Purification

MembraneMax Protein Expression Kits allow the expression and purification of soluble membrane proteins to overcome some of the hurdles preventing a deeper understanding of membrane protein structure and function. The kit combines an optimized nanolipoprotein reagent with a cell-free expression system, enabling the scalable expression and purification of up to milligram quantities of pure protein. These membrane protein–MembraneMax complexes can be used for antibody production, biophysical studies, and assay development.

Invitrogen
For information 800-955-6288
www.invitrogen.com/membranemax

Apoptosis Detection

The Guava Nexin Assay monitors the induction and progression of apoptosis quickly and accurately. The assay provides all necessary reagents in a single cocktail to minimize assay steps, assay time, and sources of contamination. The assay relies on a two-dye strategy. The dye Annexin V-PE detects the translocation of phosphatidylserine to the external surface of the membrane of apoptotic cells, an early indication of apoptosis. The assay also includes the dye 7-AAD, which is excluded from live healthy cells and early apoptotic cells, but permeates late-stage apoptotic and dead cells. The two dyes allow the user to discern between early- and late-stage apoptosis cells quickly and with minimal sources of error.

Guava Technologies
For information +44-1780-764390
www.guavatechnologies.com

Small Animal Imaging

The Leica FCM1000 is for real-time in vivo and in situ imaging of fluorescence in mice and rats. This fibered, microendoscope allows researchers to image deep brain events, peripheral nerves, and angiogenesis in a minimally invasive fashion with cellular resolution. The instrument's microprobes are designed to access virtually anywhere inside the living animal. By simply contacting the tissue of interest, the user can generate high-speed recordings of cellular or vascular events. The flexibility and minute diameters of the Leica Fibered and Miniaturized Microprobes enable endoscopic access to the living animal, with minimal animal preparation. The Leica FCM1000 allows real-time (12 frames/second), in vivo observation with minimal disturbance of the process under investigation. Physiological events or pathophysiological processes can be imaged and quantified in real time when they occur. The limited invasiveness of the technique allows repeated acquisitions on the same animal with intervals of days or weeks between time points.

Leica
For information 800-248-0123
www.leica-microsystems.com

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/products/newproducts.dtl for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availabilty of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

LEADERSHIP, STABILITY, AND SOCIAL RESPONSIBILITY

This year's survey of top employers features a familiar winner. The survey also points out the critical importance of flexible, innovative leadership and responsible behavior toward employees, customers, and potential customers in maintaining a company's high reputation. **By Peter Gwynne**

In this year's *Science* survey of top employers the industry sees the return of a past winner. For the sixth time in the survey's seven years, respondents have judged Genentech to be the best of the best. After taking top position in the first five surveys, the South San Francisco biotechnology company placed second last year, but this year is back at No. 1.

The survey reveals two other interesting developments. For the first time a company outside the mainstream pharmaceutical and biopharmaceutical business has reached the pinnacle of top employers. Survey respondents voted into second place Monsanto Company, a company that applies biotechnology as well as traditional chemistry to create and manufacture its agricultural products. In addition, this year's list of top employers has a strong international flavor. Four of the top 10—Boehringer Ingelheim, Roche Pharmaceuticals, EMD Serono, and Millennium: The Takeda Oncology Company—and eight of the top 20 are head-quartered outside the United States. Firms based in the United States—Genzyme Corp., Schering-Plough Corp., Gilead Sciences, and Eli Lilly and Company—round out the top 10.

The survey also revealed greater emphasis on the leadership qualities of companies judged to be the best employers. While *being an innovative leader in the industry* remains the major driving characteristic of top employers (see figure on p. 288), respondents nominated *having top leadership that successfully makes changes needed to keep the organization moving in the right direction* as one of the half-dozen most important drivers; that driver did not make last year's top six. Respondents also use social criteria in judging the top employers. *Being socially responsible, having loyal employees, treating employees with respect,* and *having work culture values aligned with employees' personal values* make up the rest of the six most important drivers.

Interviews with representatives of highly ranked companies reveal the basic ingredient that marks a top employer: leadership that ensures an atmosphere of stability for employees and customers in a rapidly changing business environment. In most cases, that approach stems from the corporation's basic values (see p. 292).

"We have a core of values that we use to run the company," says **Richard Scheller,** Genentech's executive vice president of research and chief scientific officer. "We don't deviate from these values when it might be better for the bottom line or opportunistic." **Steve Padgette,** vice president of biotechnology at Monsanto, regards the consistent relationship between R&D and commercial products as the key to success. "At the core, it's the company we are, being leaders in innovation," he says.

Continuity and Change

Richard Gregory, senior vice president and head of research at Genzyme, sees maintaining a corporate culture rooted in consistent values as the basis of his company's third-place ranking. "We have maintained the values of commitment to innovation and focus on unmet needs, even as we have doubled in size over the past six or seven years," he says. And fourth-placed Boehringer Ingelheim, founded in 1865, has successfully negotiated the chasm between continuity and change. "We work very hard to be ahead of the change and

> **❝**Science drives what we do here. We follow what the science shows us will be the best for the patients.**❞**

UPCOMING FEATURES

Top Twenty Employers

1	2	Genentech (South San Francisco, CA)	Innovative leader in the industry	Loyal employees	Treats employees with respect
2	8	Monsanto Company (Creve Coeur, MO)	Innovative leader in the industry	Loyal employees	Leadership moving company in right direction
3	5	Genzyme Corp. (Cambridge, MA)	Innovative leader in the industry	Socially responsible	Loyal employees
4	1	Boehringer Ingelheim (Ingelheim, Germany)	Loyal employees	Treats employees with respect	Work and personal values are aligned
5	4	Schering-Plough Corporation (Kenilworth, NJ)	Socially responsible	Loyal employees	Innovative leader in the industry
6	18	Roche (Basel, Switzerland)	Innovative leader in the industry	Loyal employees	Leadership moving company in right direction
7	–	EMD Serono, Inc. (Rockland, MA)	Socially responsible	Work and personal values are aligned	Loyal employees
8	13	Millennium: The Takeda Oncology Co. (Cambridge, MA)	Innovative leader in the industry	Treats employees with respect	Socially responsible
9	16	Gilead Sciences (Foster City, CA)	Innovative leader in the industry	Leadership moving company in right direction	Socially responsible
10	11	Eli Lily and Company (Indianapolis, IN)	Socially responsible	Loyal employees	Innovative leader in the industry
11	6	Novartis (Basel, Switzerland)	Innovative leader in the industry	Socially responsible	Leadership moving company in right direction
12	9	AstraZeneca PLC (London, UK)	Innovative leader in the industry	Socially responsible	Loyal employees
13	15	Wyeth (Madison, NJ)	Innovative leader in the industry	Loyal employees	Socially responsible
14	14	Merck & Co., Inc. (Whitehouse Station, NJ)	Innovative leader in the industry	Loyal employees	Socially responsible
15	10	Johnson & Johnson (New Brunswick, NJ)	Socially responsible	Loyal employees	Innovative leader in the industry
16	–	sanofi-aventis (Paris, France)	Loyal employees	Treats employees with respect	Socially responsible
17	7	DuPont (Wilmington, DE)	Socially responsible	Loyal employees	Innovative leader in the industry
18	17	Biogen Idec (Cambridge, MA)	Innovative leader in the industry	Socially responsible	Treats employees with respect
19	12	GlaxoSmithKline (London, UK)	Innovative leader in the industry	Socially responsible	Loyal employees
20	3	Amgen (Thousand Oaks, CA)	Innovative leader in the industry	Socially responsible	Treats employees with respect

The 20 companies with the best reputations as employers, according to respondents in the 2008 survey undertaken for the *Science* Business Office, and their three highest-scoring drivers. The companies without a 2007 rank did not receive enough mentions to qualify during the 2007 survey.

to have the flexibility and willingness to look at things differently as time passes," says **David Nurnberger,** senior vice president for human resources.

The top companies have proved successful in two other respects. They have avoided the negative perceptions that, according to survey respondents, affect much of the life science industry. "If you live by fundamental core values and apply them to every aspect of your mission," Scheller says, "you tend not to have those problems." At the same time, top employers have emphasized and benefited from the types of achievement that provide positive views of the field. "We are flexible enough to apply new technologies, an example being our investment a few years ago in biopharmaceutical production," says **Gerd Schnorrenberg,** vice president in research at Boehringer Ingelheim in Germany. **Steve Gansler,** senior vice president, human resources at Millennium, echoes that view. "Millennium was founded based on innovation," he explains. "That's something that has continued."

Science's Business Office commissioned Senn-Delaney Culture Diagnostics & Measurement to conduct the web-based survey. As it did last year, Senn-Delaney used two approaches to solicit respondents. In one, e-mail invitations went out to members of AAAS, registrants with *Science* Careers, and visitors to *Science*'s website who had registered with AAAS. In addition, Senn-Delaney sent an e-mail blast to human resource contacts at industry firms pulled from the AAAS/*Science* Careers sales database. A total of 3,990 respondents completed the survey between April 17 and May 11 of this year. The table on page 286 shows the sample's demographics.

Best, Average, and Worst

Respondents were asked to list companies, including their own, that they regarded as the best, average, and worst employers. Sixty-two percent chose as best the company that employed them and 13 percent had past connections with that company. Survey takers then rated the companies they had chosen on 23 driving continued »

Demographics

Gender:
54% male, 46% female

Experience:
62% have 10 or more years work experience

Career Status:
74% report that they have not yet reached the peaks of their careers

Company Type:
38% biotech, 34% pharma, 14% biopharma; more than four out of five work in private industry

Geography:
12% in continental Europe; 80% from North America

characteristics. Senn-Delaney applied a statistical process to those ratings to create a unique ranking score for each company rated by at least 30 respondents.

This year's top 10 employers fall naturally into three groups. Genentech, Monsanto, Genzyme, and Boehringer Ingelheim occupy the top tier, with ranking scores over 90 out of 100. Next in order come Schering-Plough, Roche, and EMD Serono, garnering between 71 and 89 ranking scores. And Millennium, Gilead, and Eli Lilly complete the top 10 list with scores up to 70.

Not surprisingly, the top 10 employers scored well on several of the top six drivers. Genentech, for example, earned top grades or close to top grades on all six. Scheller puts that down to the company's dedication to doing excellent research. "Science drives what we do here," he explains. "Decisions are made by scientific process and based on scientific data. We follow what the science shows us will be the best for the patients."

Sherri Brown, Monsanto's vice president of chemistry, attributes her company's high grades to leadership in a similar way. "We are a technology-driven company," she says. "Our products come from innovation. So it's important for our scientists to do cutting-edge work." Genzyme also cites innovation to explain its high scores in innovative leadership, social responsibility, and loyal employees. "Innovation is risky," Gregory says. "We're a company with a relatively high tolerance for risk."

Innovation and Cultural Comfort

Innovation emerges as the major theme of almost every company in the top 10. "We continually strive to create an environment where our scientists have the freedom and flexibility to innovate," says **Bernhard Kirschbaum,** executive vice president for research at Merck Serono, which operates in North America as EMD Serono. "We place tremendous value on integrity and innovation," says Millennium's Gansler. "Innovation is fundamental to all aspects of our business," adds **Norbert Bischofberger,** Gilead's executive vice president, research and development, and chief scientific officer.

Cultural comfort between company and employees provides the other main set of drivers evident in top employers. Boehringer Ingelheim, for example, sees stability as the key to its high rating in employee-related drivers. "It's really important that we are privately held and don't have to worry about the stock price," Nurnberger says. "It enables us to provide a more stable environment for our

employees for building strong teamwork within the organization."

Steven Paul, executive vice president of science and technology and head of R&D, explains the importance of employee loyalty to Lilly. "We like to say we hire people to work for their entire career," he says. "This is increasingly challenging as priorities and programs change and as we restructure the way we pursue biopharmaceutical R&D. However, we believe we've done it in a way that values our scientists and their critical contributions to the success of our company." Much the same applies to sanofi-aventis, a company that has consistently ranked highly, often in the top 20. "Respect and solidarity are part of our defined values," explains global head of R&D **Marc Cluzel.** "Working in R&D needs motivation as well as excellent science. It's really critical to ensure that your people are motivated."

A critical part of motivating scientists involves exposing them to tough problems. "Our scientists want challenging, innovative work," Brown says. Paul sees the existence of high-quality science as a key recruiting tool for Lilly. "I think that scientific reputation is the most likely driver of the top employer ratings," he says. "Scientists from the outside recognize the excellence of Lilly's basic scientific, as well as clinical, research." Scientific excellence can also facilitate collaborations with universities and other organizations. "If you want to attract employees and collaborate with researchers outside, you need to be attractive scientifically," Cluzel explains.

Tailoring Treatments

A particularly attractive, and demanding, area of research focuses on tailoring treatments to individual patients. "The thing we're particularly proud of is our innovation in personalized medicine," Genentech's Scheller says. "We're developing diagnostics for every drug in our pipeline." EMD Serono, Kirschbaum says, "focuses on the development of targeted cancer therapies on three therapeutic platforms: targeting the tumor cell, the tumor environment, and the immune system." At Lilly, adds Paul, "We have the overriding mantra of tailored therapies—the right drug for the right patient at the right time for the right duration. We try very early on to identify subgroups that will do best on specific drugs."

One member of the top 10 employers shows that challenging research in biotechnology is not restricted to the pharma and biopharma industries. "No longer is agricultural research following behind what happens in pharma," Monsanto's Brown declares. "The convergence of technologies applying to agriculture is fabulous," adds her colleague Padgette. "You can walk into a lot of our laboratories and the work is the same as in a pharmaceutical company."

Several other themes emerged within the contexts of the key drivers of leadership, social responsibility, and respect for employees. "It's my job to keep science as a fun enterprise for people here at Genentech," says Scheller. "Scientists love getting up in the morning, thinking about data, and trying to solve scientific problems that help patients." At EMD Serono, says Kirschbaum, "We strive to create an environment that fosters growth and professional development continued »

Driving Characteristics of Top Employers

2008	2007
1. Innovative leader in the industry	1. Innovative leader in the industry
2. Socially responsible	2. Treats employees with respect
3. Loyal employees	3. Work and personal values are aligned
4. Treats employees with respect	4. Loyal employees
5. Leadership moving company in right direction	5. Socially responsible
6. Work and personal values are aligned	6. Clear vision toward the future

Black type indicates the characteristics in common for the past two years.

and a place where intellectually curious people can thrive and grow in new experiences." Lilly also puts high value on its scientific tasks. "If you're doing innovative R&D, a lot of great science will come out of it," Paul says. "Our job is to create medications. But the science is absolutely essential."

Tangible Results

Scientists also like to see the tangible results of their endeavors. "When scientists face the choice of an academic career or industry, many of those choosing industry do so because they want to make a difference that will improve people's lives," Gregory says. "That is a powerful incentive to move to industry and stay there." Gansler agrees. "Everybody wants to have the opportunity to make a difference and have an impact," he says. That mode of thinking applies beyond medical products. "We have an innovation and scientific excellence focus, but also a product focus," Monsanto's Padgette says. "There's a connection between our science and our products—between R&D and our commercial plans—that makes a very large difference."

Equally intriguing to scientists is the opportunity to break new ground. "We are looking for unmet therapeutic needs," says Boehringer Ingelheim's Schnorrenberg. "In an interdisciplinary approach, we look at how we can address them. We recently introduced the first oral thrombolytic drug on the market in more than two decades. And we have several cancer and metabolic projects in advanced clinical investigation." Gilead takes a similar approach. "We ensure our continued success as a top employer by remaining focused on what we do best—developing innovative new medications for life-threatening diseases and getting them to patients who need them," Bischofberger says.

Humanitarian work plays a strong role in the perception that companies act in a socially responsible way. Genzyme, for example, is committed to developing innovative therapies for diseases such as malaria and sleeping sickness that have largely disappeared in the industrial nations but affect millions in Third World countries. "We do not intend these as future profit makers, but we still work on them," Gregory says. "Our employees are tremendously excited by this." Social responsibility is part of the culture at other companies, too. "One reason for the industry's bad image is the feeling that we are interested only in people able to pay a lot of money for our products," Cluzel points out. "Here at sanofi-aventis, we are working to find ways to provide drugs to patients who can't afford them."

The International Perspective

The number of top employers headquartered outside North America raises the issue of the importance of a worldwide presence for modern life science companies. "It is extremely important for a pharmaceutical company to be global, mainly to be able to access markets and attract skilled employees," Schnorrenberg says. "At Boehringer Ingelheim we operate globally for both marketing operations and research centers." The company also recruits worldwide, as evidenced by the 17 different nationalities represented at its headquarters.

Life science companies based in the United States have their own positive take on a global perspective. "Ours is a business that requires maximum scientific and intellectual input to be successful," Paul says. "Lilly has been on the cutting edge of establishing research centers outside the United States, particularly in Asia." Adds Bischofberger: "At Gilead we saw value in retaining control of our products outside of North America. We have invested in setting up or expanding existing operations in Canada, Europe, and Australia."

Two of the top 10 companies have a particular interest in globalism, having been involved in recent international acquisitions. Merck KGaA, a Darmstadt, Germany, company founded in 1668 by the family that makes up a majority of its present ownership, acquired Serono SA in January 2007, creating a division of the parent company named Merck Serono. Due to trademark restrictions, this division operates in the United States and Canada as EMD Serono (ranked 7th) to distinguish it from Merck & Co. (ranked 14th), a separate and independent company. "Our centers of excellence located in the United States, Switzerland, Germany, and Italy provide greater synergy and collaboration in effective networks," Kirschbaum says. Meanwhile, Millennium was acquired by Japanese pharma Takeda Pharmaceutical Company in May of this year. "It's very attractive to be part of an organization that is global in nature," Gansler says. "We have ready-made markets and a ready-made structure to sell our products on a worldwide basis."

Negatives and Positives

The life science industry has had its share of problems in recent years. In many cases, negative perceptions of the industry exceed the positive characterizations. Respondents in this year's survey pointed to five main causes of the field's less than favorable reputation: drug and product recalls such as the withdrawal of Avandia; safety issues such as the discovery of problems with raw material from China used in medical products; scandals, including evidence that pharmaceutical companies have failed to release data from trials whose results cast doubts on their drugs' safety and efficacy; lawsuits brought against companies that failed to warn patients of problems with their products; and ethical issues such as kickbacks for physicians promoting specific medications. "A lot of people appreciate what the industry is doing for patients," Scheller says. "But the industry has got itself into a little bit of a difficult situation because of those issues." continued »

You can find an expanded version of this feature by going to: dx.doi.org/10.1126/science.opms.r0800061

Survey respondents had no doubt about the treatment for the industry's problems: information sharing. "Be open and transparent," one recommended. "Communicate to the public the worth of what companies are doing without overhyping and driving up expectations," another suggested. That message has plainly got through to the top employers, exemplified by sanofi-aventis: "We understand that we have to fully articulate the economic and social value of our products," says Cluzel.

Perceptions of the industry aren't entirely negative. Respondents noted the positive impact of general advances in health care. These include the ability to bring new, innovative products to the market; progress in oncology, such as the approval of Genentech's drug Avastin for breast cancer and advances in the use of gene therapy, despite some setbacks; continuing progress in research on stem cells; the emergence of treatments for previously untreated diseases; and solutions for food and fuel shortages, such as biocrops and biofuels.

Putting the Patient First

Bischofberger points to the impact of successful research on one key disease. "The antiretroviral treatments now available mean it is possible for a young person diagnosed with HIV today to have an estimated survival rate of approximately 35 years," he says. Like Gilead, other top employers accentuate the positive aspects of the life science industry by performing their R&D in entirely new areas. "We're very progressive in terms of our innovative approach to health care," Gregory of Genzyme says. "We're definitely trying not to be a me-too company."

Lilly's Paul points out the fundamental philosophy that garners a positive perception of the industry. "Put the patient first across the entire value chain, from R&D to marketing," he advises. "If you come to work every day to do what's right for patients and to develop drugs with that in mind, you'll be successful."

That approach reflects the attitude of survey takers to another aspect of the industry: the field's attraction to scientists. The main advantage of working in the life science industry, respondents reported, is the chance to make a difference, by performing work that brings the reward of helping people. "One of the most important driving forces for individuals is being empowered to take responsibility for projects on an individual basis as part of a team," Boehringer Ingelheim's Schnorrenberg says. "We are striving to encourage scientists and coach them."

Working in the life science industry is not without drawbacks. More than a quarter of the survey's respondents reported that they are fairly likely, very likely, or extremely likely to seek a different job. Most quoted job insecurity as their reason. That stems in large part from corporate mergers and acquisitions, restructuring, and outsourcing.

Several top employers have a policy of minimizing those causes of uncertainty. "We are very conservative in our hiring," Genentech's

Bubble width = Respect

Comparison of the top 10 companies on the basis of 3 of the top drivers (scored out of 100): *Is an innovative leader in the industry* (x-axis), *Has top leadership that successfully makes changes needed to keep the organization moving in the right direction* (y-axis), and *Treats employees with respect* (bubble width).

Scheller says. "The good part is that we won't have large layoffs. The downside is that, when a clinical trial works and promises large revenues, we have to stimulate our research very quickly." Lilly has purposely avoided major mergers and acquisitions. "We decided that we didn't need to become too much larger through merging with a large company," Paul says. "We will continue to grow organically through small mergers and acquisitions."

A Stable Environment

Boehringer Ingelheim points out that its private ownership allows it to make long-term decisions without worrying about the next-quarter impact. "Many candidates I interview for jobs have been part of companies that have gone through mergers and acquisitions," Nurnberger says. "They are always interested in Boehringer Ingelheim because they see a stable environment."

Life science companies can't always avoid some insecurity stimulated by change. For them, the guiding principle is to minimize personnel dislocations. "Our executive team has articulated a company strategy that puts acquisitions into the context of our mission," Monsanto's Padgette says. At Genzyme, Gregory adds, "We have done a number of mergers and acquisitions, but we have maintained our culture of innovation and collaboration. As we add new talent to our work force, we allow our people to stay with projects or to move on to suitable new projects. We are all working together toward the larger goal of making sure every patient who needs therapy is treated."

EMD Serono and Millennium have the most recent experience of change, having undergone acquisition early in 2007 and in May 2008, respectively. "The acquisition [of Serono] allowed us to redefine processes, examine our pipeline and research initiatives, and define our core therapeutic areas," Kirschbaum says. "As a result, we have a strong, attractive pipeline strategically aligned within the organization's core objectives." Before Takeda acquired Millennium, both companies made sure that they were, in Gansler's words, "a great cultural fit." Beyond that, he continues, "Takeda made it clear that the transaction was about growth and not synergies. They also made it clear how they valued Millennium's employees, putting retention bonuses in place. So the issue of job security didn't arise."

Gilead's Bischofberger defines the tightrope that corporate executives must negotiate as they plan the future. "This is certainly a dynamic and competitive industry, and change—good, bad, and in between—is constantly taking place and always will be taking place," he says. "What is a constant for us is our focus on developing and delivering innovative treatments for unmet medical needs and making sure that our employees have a work environment that inspires a commitment to this mission."

A former science editor of Newsweek, *Peter Gwynne writes about science and technology from his base on Cape Cod, Massachusetts, USA.*